

RICE UNIVERSITY

**A Flexible Framework for the Examination of
Production, Measurement, and Contracts
in the Face of Moral Hazard**

by

Jonathan David Bonham

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

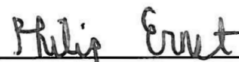
APPROVED, THESIS COMMITTEE:



Thomas Hemmer, Chair
Houston Endowment Professor of Accounting



K. Ramesh
Herbert S. Autrey Professor of Accounting



Philip Ernst
Assistant Professor of Statistics

Houston, Texas

April, 2017

A Flexible Framework for the Examination of Production, Measurement, and Contracts in the Face of Moral Hazard

Jonathan David Bonham

Thesis: Doctor of Philosophy
Accounting, Jones Graduate School of Business
Rice University, Houston, Texas (April 2017)

Abstract

I develop a framework to examine the effect of measurement on productive activity in the face of moral hazard. I allow an agent intricate control over the stochastic value of a firm's assets, and he is compensated based on a report produced by an accounting system that admits a large class of bias- and timing-oriented accounting measurement rules.

When measurement error is unavoidable but is treated to address the moral hazard problem, (i) the fundamental earnings distribution develops asymmetric tails and discontinuities at predictable thresholds, (ii) measurement rules develop all-or-nothing recognition properties and are rarely unconditionally biased, and (iii) the contract develops caps, floors, and hurdle bonuses at predictable thresholds.

In contrast, when measurement error can be reduced by delaying measurement until uncertainty has been resolved, historical cost accounting is unambiguously optimal in curtailing moral hazard. However, I show that an accounting regulator with alternative objectives can influence economic activity by mandating timely measurement. Specifically, I show that timely loss recognition induces firms that are more (less) averse to downside risk to contract for riskier (less risky) actions.

Finally, I show that first best actions are implementable in my setting via a two-wage penalty contract only if the measurement rule is extremely noisy and unconditionally conservative. Furthermore, the agent charges a negligible risk premium if he is sufficiently optimistic about the odds of avoiding a penalty-triggering earnings report. In other words, unconditionally conservative measurement can disable moral hazard when the agent is optimistic.

Acknowledgements

I express my sincere gratitude to my advisor, Thomas Hemmer, for his constant support, insight, encouragement, and guidance throughout my doctoral studies. He has defended the integrity of my education heroically, has provided every opportunity to gain exposure to thought leaders in the field, and has more than delivered on every explicit and implicit commitment that he's made to me. I couldn't ask for a better advisor or mentor.

I would also like to thank the members of my committee, K. Ramesh and Philip Ernst, for their hard questions, pivotal advice, and scholarly examples. And many thanks to the accounting, finance, and economics faculty who have contributed to my education, and to my fellow doctoral students who have shared many long and memorable hours of study and discussion.

And finally, I offer a very special thanks to my wife, Lindsey, for her constant support and encouragement, and to my parents, David and Cindy, who instilled in me a love for learning at an early age.

Contents

Abstract	ii
Acknowledgements	iii
Prologue: Simplification by Generalization	v
1 Stewardship-Oriented Production, Measurement, and Contracts	1
1.1 A general and flexible framework for the study of moral hazard	7
1.2 Incentive-compatible production, measurement, and contracts	11
1.3 Stewardship-oriented production, measurement, and contracts	16
2 Standardizing Production through Uniform Reporting	27
2.1 Equating timeliness with measurement error in a multiperiod model	29
2.2 Influencing downside risk through timely loss recognition	31
3 An Optimistic Case for Pessimistic Measurement	34
3.1 A rank dependent expected utility representation of optimism	37
3.2 Penalty contracts with negligible risk premia	41
Appendix A Proofs	50
Appendix B Figures	63

Prologue: Simplification by Generalization

In this collection of essays I present a general framework that relaxes many of the restrictions imposed in the extant moral hazard literature, and I demonstrate that relaxing these restrictions produces novel and empirically relevant predictions. The classical moral hazard problem is characterized by a principal who employs an agent to provide some unobservable productive input. Since this input, or action, is unobservable, the agent is inclined to act in his own best interests unless he is incentivized to do otherwise. Therefore, the principal's problem is to select (i) an *action* to incentivize, (ii) a *performance measure* that is informative about the action actually taken, and (iii) a *contract*, written on the performance measure, that induces the agent to take the desired action.

In an attempt to solve this complex problem, academic researchers have tended to place restrictions on one or more of the principal's choice variables. The most common approach is to assume that the agent's action *parametrically* influences the distribution of economic profit consumed by the principal. A common representation is to allow the agent to choose an action $a \in [\underline{a}, \bar{a}]$, where a affects the central tendency of the distribution of profit $f(\pi|a)$.

While this approach simplifies the agent's action space, it creates an additional technical hurdle. Namely, it implies that there are many contracts that induce the agent to take any given action a . The principal's (and, therefore, the researcher's) problem is thus to identify the *best* contract to offer among the set of incentive compatible contracts. Generally speaking, nailing down the properties of the optimal contract tends to burn up most of the researcher's mathematical gunpowder. Once this has been done, very little can be said about the properties

of optimal actions, which, in all fairness, aren't all that interesting in the first place given the restriction on the agent's action space.

In this work I employ an assumption initially introduced to the literature by Holmstrom and Milgrom (1987), and allow the agent *nonparametric* control over the distribution of economic profits. By allowing the agent to choose the probability of every outcome, (i) the number of incentive compatible, individually rational contracts is reduced to one, and (ii) the agent's actions become much more interesting. By taking this approach, I no longer need to exhaust my degrees of freedom in nailing down the properties of the optimal contract; the unique implementing contract in my setting is very simple. Instead, I can focus my attention on the properties of optimal actions, which can now include empirically relevant behavior such as asset substitution and real activities manipulation. In other words, a more general action space actually makes the contracting problem *easier* to solve.

I begin in Chapter 1 by developing the framework and by characterizing productive actions, measurement rules, and contracts that solve the moral hazard problem when performance measures are subject to measurement error. First, I show that the optimal action induces a fundamental earnings distribution with asymmetric tails and discontinuities at predictable thresholds, consistent with empirical evidence. Second, I show that optimal measurement rules possess all-or-nothing recognition properties – consistent with the recognition criteria for contingent liabilities, leases, and certain allowance accounts – and are rarely unconditionally biased. Finally, I show that the optimal earnings-based contract develops caps, floors, and hurdle bonuses at predictable thresholds, which is also consistent with practice and notably inconsistent with nearly all prior theoretical characterizations of optimal contracts.

In Chapter 2 I assume that the contracting parties take the measurement rule as exogenous, chosen by an accounting regulator. I derive a formal link between timeliness and measurement error in this setting, and I show that timely loss recognition can either increase or decrease downside risk depending on cross-sectional firm characteristics such as liquidity. I also suggest

that uniform accounting standards can actually cause firms' production and risk profiles to become more similar. While my work in this chapter is preliminary, I believe there is great potential for better understanding the impact of accounting regulation on individual contracts and aggregate production.

Finally, in Chapter 3 I provide sufficient conditions for severe noise and conservative bias to be strictly desirable in the face of moral hazard. I show that first best actions are implementable in my setting via a penalty contract designed after Mirrlees (1974) only if the measurement system is extremely noisy and unconditionally conservative. Furthermore, the agent charges a negligible risk premium if he is sufficiently optimistic about the odds of avoiding a penalty-triggering earnings report. The model predicts that optimistic managers receive less variable compensation and demand lower risk premia than do neutral or pessimistic managers. To the extent that innovative managers are more optimistic and are better able to augment asset value, the model also rationalizes unconditionally conservative measurement of research and development and other innovation-dependent intangibles.

Chapter 1

A Flexible Framework for the Examination of Stewardship-Oriented Production, Measurement, and Contracts

“A [theory] is important if it ‘explains’ much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone.”

— Henry Friedman (1953)

In this chapter I develop a general representation of the moral hazard problem that places very few restrictions on productive activity, accounting measurement, and contractual form. I demonstrate that this formulation of moral hazard is capable of rationalizing many empirical regularities that have long interested and puzzled accounting researchers. Specifically, the model explains (i) asymmetric earnings distributions with discontinuities at predictable thresholds, (ii) all-or-nothing recognition criteria and conditionally-biased measurement rules, and (iii) contracts that exhibit caps, floors, and hurdle bonuses. While several of these regularities can be explained by alternative idiosyncratic theories, this chapter presents a holistic theory of moral hazard that is capable of explaining “much by little.”

The classical moral hazard problem is characterized by a principal who employs an agent to provide some personally costly, unobservable, but valuable productive input, or action. Since

the action is unobservable, the agent is inclined to take an action that is suboptimal from the principal’s perspective unless he is incentivized to do otherwise. Therefore, the principal’s problem is to select (i) an *action* to incentivize, (ii) a *performance measure* that is informative about the action actually taken, and (iii) a *contract*, written on the performance measure, that induces the agent to take the desired action.

Because the problem consists of three complex choice variables, a general mathematical solution has long evaded academic researchers. The universally-applied remedy is to curtail the problem’s complexity by placing exogenous restrictions on the set of feasible actions, the properties of performance measures, and/or the contractual form. For example, Holmström (1979) abstracts away from the selection of a performance measure and restricts the set of actions to those that satisfy the *Monotone Likelihood Ratio Condition*. Research based on Gigler and Hemmer (2001) considers properties of the performance measure such as conservatism, but typically restricts attention to a binary action space (e.g., high versus low effort) to do so. The “LEN” literature imposes strong exogenous assumptions on all three choice variables: contracts are restricted to be linear in normally-distributed performance measures whose means – and means alone – are chosen by the agent.

In this work I leverage an assumption introduced by Holmstrom and Milgrom (1987) that allows me to simultaneously characterize optimal actions, performance measures, and contracts while avoiding severe exogenous restrictions on these choice variables. Specifically, I allow the agent to independently control all moments of the production function by letting him directly choose the probability of each outcome. In more applied terms, the agent has power over the firm’s strategy, projects, product mix, risk profile, growth rate, and indeed its most fundamental characteristics.¹ This is in sharp contrast to the majority of the literature which

¹While Holmstrom and Milgrom point out that their approach offers opportunities to better understand the selection of these characteristics, their focus lies primarily on identifying a setting that produces tractable contracts. As they push their model to its continuous time limit, the agent’s intricate control over the distribution of firm output disappears and he is left controlling only the mean of a normal distribution.

allows the agent independent control of the mean, and sometimes the variance, of firm output.² By expanding the agent’s control over production, it turns out that the moral hazard problem actually becomes *easier* to solve.

The second ingredient in my framework is a novel analytical representation of an accounting system that admits a large class of measurement alternatives. The system is characterized by two parameters per fundamental outcome, each respectively specifying the probability of aggressive and conservative measurement given that outcome. The sum of these parameters determines the degree of noise or measurement error in the accounting system, which I here assume is prohibitively costly to eliminate. More importantly, the realization rule, fair value, unconditional conservatism, conditional conservatism, unconditional aggression, and conditional aggression are each captured by appropriate specifications of the aggressive and conservative parameters.

The model’s solution yields two sets of empirically descriptive predictions. The first set of predictions is driven by measurement error, which increases dispersion in earnings. This dispersion makes extreme earnings realizations more likely, thereby exposing the agent to additional compensation risk, all else equal. In response, the principal shields the agent from this excess risk by introducing caps and floors to the contract. Moreover, the capped and floored regions become larger as the agent’s risk aversion increases or as measurement error becomes more severe. Murphy (1999, 2013) documents that caps and floors are the empirical norm in accounting-based bonus plans, but they are notably inconsistent with nearly all prior theoretical

²Very few moral hazard studies allow the rich set of distributions afforded by Holmstrom and Milgrom’s nonparametric action assumption. Two notable exceptions are Hellwig (2007) and Bertomeu (2008), whose findings uncover a potential deficiency in models that only allow the agent control over a single central parameter. That is, the principal’s primary concern as he moves from a first best to a second best world is how to trade off actions that are productively efficient with actions that expose the agent to less risk. Since risk exposure depends on both the implementing contract *and* the implemented distribution, an important lever that the principal would like to pull is missing in models where actions are ranked only on average productivity.

characterizations of optimal contracts.³ In fact, Murphy and Jensen (2011) call for practitioners to eliminate caps and floors from compensation contracts since they appear suboptimal based on existing theories. My model indicates that such a prescription may be premature.

While caps and floors shield the agent from excessive measurement risk, they also decrease the agent's incentive to take actions that increase upside risk and decrease downside risk. Such actions lead to asymmetry in the tails of the fundamental earnings distribution. Prior empirical studies suggest that the observed direction of asymmetry depends heavily on the specification (e.g., see Panels A versus B of Fig. 1 in Durtschi and Easton 2005, and Figure 2 versus Figure 4 in Beaver, McNichols, and Nelson 2007), while others have documented cross-sectional variation in earnings skewness (e.g., Gu and Wu 2003). My model indicates that a greater degree of measurement error and moral hazard is likely to promote fundamental earnings distributions with thicker lower tails and thinner upper tails.

Not only do caps and floors decrease the agent's incentives to take actions with high upside and low downside risk, but they also promote actions that are less productive on average. This is because aggressive (conservative) measurement of moderate fundamental performance fails to expose the agent to a potential payoff above (below) the cap (floor). In order to partially restore these damaged incentives, the principal chooses a measurement rule that makes the caps and floors more likely to attain; that is, he chooses a measurement rule that is aggressive over relatively high outcomes and that is conservative over relatively low outcomes. Such a measurement rule parallels "all-or-nothing" recognition criteria that are common in practice. For example, contingent liabilities are either recorded at their probable amounts or are left off the balance sheet altogether; an allowance for deferred tax assets is either booked at the asset's full value or is completely ignored; capital leases are recorded at value whereas operating leases have no place on the balance sheet.

³One partial exception is Arya, Glover, and Mittendorf (2007), who justify bonus caps, but not floors, in a setting with organizational hierarchies.

My second set of predictions hinges on the ability of the measurement system to cleanly distinguish performance that exceeds some threshold from performance that falls short. Auditing is likely to create such separation, since auditors have an increased mandate to disallow misstatements across qualitatively significant thresholds.⁴ These thresholds are likely to include zero or consensus analyst forecasts, since internal and external parties have incentives to closely scrutinize measured performance approaching these values. Moreover, certain transaction characteristics themselves point to zero as a likely candidate for such threshold precision: for example, while a contingent loss may be difficult to value, it is usually easy to distinguish from a contingent gain.

Given the existence of a threshold that separates performance into distinct ranges, the contract naturally develops a hurdle bonus at the threshold as measurement error becomes more severe. This is, again, because dispersion in earnings exposes the agent to risk that can be avoided by reducing within-range variation in the contract. The result is a discrete “jump” in the contract at the threshold. Such hurdle bonuses are consistent with Murphy (1999, 2013). Moreover, I predict that hurdle bonuses are larger when the agent is more risk averse and when measurement error is more severe, and they should arise at thresholds across which misclassification is difficult or unlikely.

In order to increase his chances of securing the hurdle bonus, the agent rationally takes actions promoting a discontinuity in the fundamental earnings distribution at the threshold. In contrast to explanations based on the manipulation of reported earnings, these discontinuities arise in my model precisely because misclassification of missed versus met targets – whether

⁴Auditing standards specify that “[a]s a result of the interaction of quantitative and qualitative considerations in materiality judgments, uncorrected misstatements of relatively small amounts could have a material effect on the financial statements....” and that “a misstatement made intentionally could be material for qualitative reasons, even if relatively small in amount” (PCAOB 2010 A7 par. 17). More specifically, the standards specify that the qualitative factors to consider in the auditor’s evaluation of materiality include “[a] misstatement that changes a loss into income or vice versa” and “[a] misstatement that has the effect of increasing management’s compensation, for example, by satisfying the requirements for the award of bonuses or other forms of incentive compensation” (PCAOB 2010 A7 par. B2).

by error or manipulation – is explicitly disallowed. Moreover, the model rationalizes a contract that promotes discontinuities, whereas the earnings management explanation fails to explain why contracts do not “fix” the deviant behavior. Furthermore, the explanations can perhaps be disentangled empirically by examining other measures of fundamental performance such as cash flows, because my prediction pertains to the distribution of fundamental earnings whereas the earnings management explanation applies to the distribution of reported earnings.

Finally, I show that if the agent’s cost function satisfies a certain robustness property, then optimal measurement rules condition the direction of bias on a single threshold. Many measurement rules applied in practice possess this property, conditional conservatism and the realization rule included. While this result depends critically on idiosyncratic properties of the agent’s cost function, it turns out that typical deviations from this case result in measurement rules whose biases are conditional on *multiple* thresholds; the optimality of an unconditionally biased measurement rule is a knife-edge case that relies on perfect alignment of the agent’s preferences and the informational environment.

This chapter makes several contributions. First, it provides a tractable method for analyzing moral hazard without placing severe restrictions on the agent’s action set, properties of the performance measure, or contractual form. Second, it introduces a novel analytical representation of an accounting system that is capable of capturing the features of many measurement rules applied in practice, and which I believe has several potential applications in other settings. Third, it contributes to the growing literature on the distributional properties of earnings by rationalizing asymmetric tails and discontinuities, and by providing testable predictions about the determinants of asymmetry. Fourth, it contributes to the literature on accounting measurement by rationalizing conditionally-biased measurement rules with all-or-nothing recognition properties. Finally, this chapter contributes to the compensation literature by rationalizing the use of bonus caps, floors, and hurdle bonuses in accounting-based bonus plans – a feature

that has received very little prior theoretical support – and by producing falsifiable predictions about the location of hurdles and the size of the incentive zone.

1.1 A general and flexible framework for the study of moral hazard

I consider a single period principal-agent model in which an agent has control over some asset owned by the principal. The agent chooses the distribution of fundamental earnings $\pi \in \{\pi_0, \pi_1, \dots, \pi_N\}$, where π_i is strictly increasing in i for $i \in \{1, \dots, N\}$. As in Holmstrom and Milgrom (1987), I allow the agent direct control over each probability p_i for $i \in \{1, \dots, N\}$, where $p_0 = 1 - \sum_{i=1}^N p_i$ ensures that the probabilities sum to one.

As this assumption plays a critical role in the upcoming results, some discussion of its appeal is warranted. One interpretation is that the agent selects multiple actions from a rich set that, when combined, interact to produce the desired distribution. As Holmstrom and Milgrom point out, even if the agent chooses a one-dimensional action such as effort, the ability to condition this action on private information received after the contract is signed expands the agent's control over the unconditional distribution. Alternatively, the agent could be choosing a sequence of contingent one-dimensional actions over time; such a contingent strategy also maps into a prior unconditional distribution at the outset. Holmstrom and Milgrom conclude that “[i]n short, we could permit rather arbitrary production and information technologies and still have the reduced form map into [this] conceptually simple structure....” Not only does the assumption seem to implicitly capture many real world situations in which the agent has some degree of productive flexibility, but it is arguably more descriptive than the more commonly employed one-dimensional action assumption.

I denote reported earnings by $x \in \{x_0, x_1, \dots, x_N\}$, and without loss of generality I assume that $x_i = \pi_i$ for all i . Departing from Holmstrom and Milgrom, I assume that the agent's utility is additively separable in wealth and effort, where the nonpecuniary cost of effort is denoted

$c(p)$ and is strictly convex in p_i for all i . I assume that $c(p)$ is twice differentiable and additively separable in the components of p , where $c_i \equiv \frac{\partial c(p)}{\partial p_i}$ and $c_{ik} \equiv \frac{\partial^2 c(p)}{\partial p_i \partial p_k} \geq 0$, with equality if $i \neq k$, for all i and k . I work in utility space and denote the agent's wage in utiles conditional on earnings realization x_i by v_i . The agent's reservation utility is denoted \bar{v} and the inverse utility function is denoted $h(v)$, which is strictly increasing and convex in v . The principal is risk neutral.

An accounting system is characterized by a set of conditional probabilities $\Pr(x|\pi)$ for all x and π . Restricting $N = 1$ and allowing the conditional probabilities to be independent, this accounting system is identical to that pioneered by Gigler and Hemmer (2001), which is often illustrated in a form similar to Figure 1.1. If I were to insist that the conditional probabilities be independent, then an extension to a model with $N > 1$ would cause the number of parameters that characterize the accounting system to grow quadratically. Precisely, there would be exactly $N(N + 1)$ independent conditional probabilities to consider. Figure 1.2 illustrates the case in which $N = 3$. The number of independent parameters suggests that tractability becomes an issue as N grows large.

I consider a more structured extension of the earnings process from Gigler and Hemmer (2001) by imposing an explicit relationship among the conditional probabilities $\Pr(x|\pi)$. Let the accounting system be characterized by the set of parameters $\Theta \equiv \{\delta_0, \dots, \delta_{N-1}\} \cup \{\gamma_1, \dots, \gamma_N\}$, and denote $\theta_i \equiv \delta_i + \gamma_i$. Noise is determined by θ_i , whereas bias is determined by δ_i and γ_i . If π_i is realized, then the accounting system measures π_i *aggressively* with probability δ_i , *conservatively* with probability γ_i , and *neutrally* or *without bias* with probability $1 - \theta_i$. Whereas the probable direction of bias is determined by γ_i and δ_i , the probable magnitude of bias is determined by the noise associated with adjacent outcomes. Specifically, if π_i is measured conservatively, then x is equal to the largest $\pi < \pi_i$ that is measured neutrally. Conversely, if

π_i is measured aggressively, then x is equal to the smallest $\pi > \pi_i$ that is measured neutrally.

Under this structure there are exactly $2N$ parameters that determine the $N(N + 1)$ conditional probabilities; the number of parameters is increasing linearly rather than quadratically in N . For mathematical expediency and with little loss of generality, I assume that $\Pr(x_i|\pi_0) = \Pr(x_0|\pi_i) = 0$ for all $i \in \{1, \dots, N\}$. Figure 1.3 illustrates this accounting system. While more complex than the binary case in Figure 1.1, it is much more manageable than the unstructured extension in Figure 1.2. Notably, there are only two instead of N measurement parameters to consider for each outcome, making it similar to the binary structure in terms of tractability.

It will be useful to distinguish among ranges of the outcome space that are *informationally distinct* from the others. I say that two adjacent ranges $\{\pi_l, \dots, \pi_{m-1}\}$ and $\{\pi_m, \dots, \pi_n\}$ are informationally distinct if a fundamental outcome in one range is never accompanied by an earnings realization in the other. That is, I exogenously restrict $\Pr(x_i|\pi_k) = 0$ if π_i and π_k are not in the same informationally distinct range. Let $J \geq 1$ be the number of informationally distinct ranges partitioning $\Pi \equiv \{\pi_1, \dots, \pi_N\}$. Then each of these J ranges can be written $\Pi^j \equiv \{\pi_{m_j}, \dots, \pi_{n_j}\}$ where $m_j = n_{j-1} + 1$. Analogously, define $X^j \equiv \{x_{m_j}, \dots, x_{n_j}\}$ for all $j \in \{1, \dots, J\}$. Finally, redefine $\gamma_{m_j} \equiv \Pr(x_{m_j}|\pi_{m_j})$ and $\delta_{n_j} \equiv \Pr(x_{n_j}|\pi_{n_j})$ for all $j \in \{1, \dots, J\}$. Figure 1.4 illustrates this modification.

These informationally distinct ranges can be given a variety of interpretations. For example, the measurement system may easily distinguish gains from losses but be unable to reliably identify the magnitude of a particular gain or loss ($J = 2$). Or it may produce reports that are precise up to the thousandth dollar, leaving only the trivial determination of rounding up, down, or to the nearest thousand to the accountant ($N = kJ$ for some $k \in \mathbb{N}$). On the opposite extreme, the accounting system may be unable to cleanly separate any range of outcomes from the others ($J = 1$), in which case the distinction is vacuous.

With this structure in place, the law of total probability allows the distribution of reported earnings to be written

$$P(x_i) \equiv \Pr(x_i|p, \Theta) = \sum_{k=m_j}^{n_j} \Pr(x_i|\pi_k)p_k \quad \text{for all } i \in \{m_j, \dots, n_j\}, \quad (1.1)$$

where $\Pr(x_i|\pi_k) = 0$ if i and k are in different informationally distinct ranges and

$$\Pr(x_i|\pi_k) = \frac{\partial P(x_i)}{\partial p_k} = \begin{cases} (1 - \theta_i)\delta_k \prod_{l=k+1}^{i-1} \theta_l & \text{if } k < i \\ (1 - \theta_i) & \text{if } k = i \\ (1 - \theta_i)\gamma_k \prod_{l=i+1}^{k-1} \theta_l & \text{if } k > i \end{cases} \quad (1.2)$$

otherwise, using the convention that $\prod_i^k(\cdot) = 1$ if $i > k$.⁵

Parallel to the definition in Gigler and Hemmer (2001), I say that the measurement rule is conservative, unbiased, or aggressive over Π^j if $E[x|x \in X^j]$ is respectively less than, equal to, or greater than $E[\pi|\pi \in \Pi^j]$. Specific measurement rules can be constructed by considering different combinations of bias over different ranges. For J informationally distinct ranges, there are 3^J possible measurement rules each specifying whether the accounting system is conservative, unbiased, or aggressive over each Π^j . The following definition characterizes a few special cases that are common in practice.

Definition 1.1. *Let gains and losses be informationally distinct, and define j° such that $\pi_{m_{j^\circ}} = 0$. Then*

1. *An accounting system applies fair value if it is unbiased over Π^j for all j .*
2. *An accounting system applies the realization rule if it is aggressive over Π^j for all $j < j^\circ$ and conservative over Π^j for all $j \geq j^\circ$.*

⁵This reported earnings distribution is equivalent over Π to the distribution of sanitized earnings in Shin (1994) if $J = 1$ and $\delta_i = 0$ for all $i \in \{1, \dots, N - 1\}$.

3. *An accounting system is conditionally conservative if it is unbiased over Π^j for all $j < j^\circ$ and conservative over Π^j for all $j \geq j^\circ$.*
4. *An accounting system is unconditionally conservative if it is conservative over all Π^j .*

To illustrate, consider an asset with some unrealized change in fundamental value as of the end of the period. The realization rule specifies that these unrealized changes in value not be recognized on the balance sheet or income statement; that is, unrealized gains are measured conservatively and unrealized losses are measured aggressively. In contrast, conditional conservatism specifies that unrealized losses, but not gains, be recognized in the current period. For example, the lower of cost or market rule for inventory valuation specifies that unrealized gains should be ignored, or treated conservatively, whereas unrealized losses should be recognized at their presumably unbiased market value.

1.2 Incentive-compatible production, measurement, and contracts

I begin by establishing the relationship among production, measurement, and contracts in this setting. A key result is that the measurement rule and the contract serve as substitutes in motivating productive activity, where the sensitivity of the agent's action to the measurement rule (contract) is increasing (decreasing) in the severity of measurement error.

The principal's program can be written

$$\begin{aligned}
 \max_{p,v} \quad & \sum_{i=0}^N \pi_i p_i - \sum_{i=0}^N h(v_i) P(x_i) \\
 \text{s.t.} \quad & \sum_{i=0}^N v_i P(x_i) - c(p) \geq \bar{v} \\
 & p \in \operatorname{argmax}_{\tilde{p}} \sum_{i=0}^N v_i \tilde{P}(x_i) - c(\tilde{p}).
 \end{aligned} \tag{1.3}$$

By the law of total probability, $P(x_i) = \sum_{k=0}^N \Pr(x_i | \pi_k) p_k$. As long as π is a sufficient statistic for p with respect to (π, x) (which is easily verified to be the case under (1.1) and (1.2)), then

$\Pr(x_i|\pi_k)$ is independent of p for all i and k , which implies that $P(x_i)$ is linear in p .⁶ Together with the convexity of $c(p)$, this implies that the agent's expected utility is strictly concave in p_k . Thus the incentive compatibility constraint's first order conditions are necessary and sufficient for an interior solution.

Notice that there are exactly N such first order conditions and one individual rationality constraint. These constraints form a system of $N + 1$ equations in $N + 1$ unknown payments, which admits a unique solution under the usual conditions. Theorem 3 of Holmstrom and Milgrom (1987) leverages this feature to characterize a unique contract that implements any given action. The following proposition is its analogue in this setting.

Proposition 1.1. *Let $P(x) \equiv \Pr(x|p, \Theta)$ characterize some earnings process in which π is a sufficient statistic for p with respect to (π, x) . For some v_0 chosen to bind the individual rationality constraint, suppose that $\{v_i\}_{i=m}^n$ is the unique solution to the following system of equations for each $(m, n) \in \{(m_1, n_1), \dots, (m_J, n_J)\}$:*

$$\begin{bmatrix} c_m \\ c_{m+1} \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} \Pr(x_m|\pi_m) & \Pr(x_{m+1}|\pi_m) & \cdots & \Pr(x_{n-1}|\pi_m) & \Pr(x_n|\pi_m) \\ \Pr(x_m|\pi_{m+1}) & \Pr(x_{m+1}|\pi_{m+1}) & \cdots & \Pr(x_{n-1}|\pi_{m+1}) & \Pr(x_n|\pi_{m+1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Pr(x_m|\pi_{n-1}) & \Pr(x_{m+1}|\pi_{n-1}) & \cdots & \Pr(x_{n-1}|\pi_{n-1}) & \Pr(x_n|\pi_{n-1}) \\ \Pr(x_m|\pi_n) & \Pr(x_{m+1}|\pi_n) & \cdots & \Pr(x_{n-1}|\pi_n) & \Pr(x_n|\pi_n) \end{bmatrix} \begin{bmatrix} v_m - v_0 \\ v_{m+1} - v_0 \\ \vdots \\ v_{n-1} - v_0 \\ v_n - v_0 \end{bmatrix} \quad (1.4)$$

Then $\{v_i\}_{i=0}^N$ is the unique wage scheme that implements p given Θ .

The system of equations in Proposition 1.1 establishes the relationship among actions, measurement, and contracts. The vector on the left hand side of (1.4) consists of marginal costs that characterize the action chosen by the manager. The square matrix on the right hand side consists of the conditional probabilities that characterize the measurement rule. Finally,

⁶By definition, π is sufficient for p with respect to (π, x) if $\Pr(x|\pi, p) = \Pr(x|\pi)$.

the vector on the right hand side consists of the earnings-contingent payments that characterize the contract. Each component is cleanly separated by (1.4).

It is clear that incentive compatible actions are determined by both the contract and the measurement rule. However, these actions are more or less sensitive to variation in the contract versus the measurement rule under certain conditions. For example, suppose that the accounting system is perfectly precise, so that $\Pr(x_i|\pi_k)$ is equal to one if $i = k$ and is equal to zero otherwise. Then the matrix in (1.4) reduces to the identity, and $c_i = v_i - v_0$ solves the system. In this case, the action taken by the manager is highly sensitive to contractual variation.

Now, for expositional expediency suppose that $n = m + 2$ and denote $k \equiv m + 1$. Moreover, invoke the earnings process specified by (1.1) and (1.2) with $\delta_m = \gamma_n = 0$. Then (1.4) reduces to

$$\begin{bmatrix} c_m \\ c_k \\ c_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \gamma_k & 1 - \theta_k & \delta_k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_m - v_0 \\ v_k - v_0 \\ v_n - v_0 \end{bmatrix}. \quad (1.5)$$

Notice that the contractual payments $v_m - v_0$ and $v_n - v_0$ completely determine c_m and c_n , but c_k becomes less sensitive to $v_k - v_0$ as measurement error increases. Specifically,

$$c_k = \gamma_k c_m + (1 - \theta_k)(v_k - v_0) + \delta_k c_n \xrightarrow{\theta_k \rightarrow \theta} \begin{cases} v_k - v_0 & \text{if } \theta = 0 \\ \gamma_k c_m + \delta_k c_n & \text{if } \theta = 1. \end{cases}$$

Holding c_m and c_n fixed, c_k becomes less sensitive to variation in $v_k - v_0$ and more sensitive to variation in (γ_k, δ_k) as θ_k increases. In the limit, c_k depends completely on the measurement rule and is unaffected by $v_k - v_0$. Even for $\theta_k < 1$, the contract's effective control over productive activity wanes as θ_k gets large. To illustrate, if the principal wishes to implement an action

satisfying $c_k \neq \gamma_k c_m + \delta_k c_n$ while holding the measurement rule fixed, then he must choose $v_k \rightarrow \pm\infty$ as $\theta_k \rightarrow 1$. Diminishing marginal utility of wealth renders $v_k \rightarrow \infty$ especially expensive for the principal, and for utility functions that are bounded above (such as the negative exponential) such wages are infeasible. Moreover, limited liability precludes any wage satisfying $v_k \rightarrow -\infty$ for a large class of utility functions.

Notice from (1.2) that $\Pr(x_k|\pi_i)$ is continuous and monotonic in θ_l for all i, k and l . While $\theta_i \in (0, 1)$ is of primary interest, continuity and monotonicity allow me to gain a clear picture of the role of measurement error on optimal actions, measurement rules, and contracts by simply considering the more tractable endpoints. It is therefore with little, if any, loss of generality that I confine attention to the extreme cases in which $\theta_i = 0$ and $\theta_i = 1$ for all $i \in \{m_j, \dots, n_j\}$. For the latter case, as illustrated in Figure 1.5, the only possible earnings realizations in X^j are x_{m_j} and x_{n_j} , which implies that the only contractual payments that are relevant for the agent's action are $v_{m_j} - v_0$ and $v_{n_j} - v_0$.

Under this assumption, the relative importance of the measurement rule versus the contract is varied by changing the size of each informationally distinct range. If $n_j = m_j + 2$, then the contract determines two thirds of the agent's chosen probabilities (c_{n_j} and c_{m_j}), whereas the measurement rule determines one third of the agent's chosen probabilities (c_{k_j}). On the other hand, if $J = 1$ then the implemented distribution is maximally dependent on the measurement rule: the contract determines only two of the agent's chosen probabilities (c_1 and c_N) whereas the measurement rule determines all other productive activity (c_2, c_3, \dots, c_{N-1}). The following proposition formalizes this statement.

Proposition 1.2. *Let $P(x) \equiv \Pr(x|p, \Theta)$ be the earnings process characterized by (1.1) and (1.2). For each $(m, n) \in \{(m_1, n_1), \dots, (m_J, n_J)\}$, suppose that $\theta_i = 1$ for all $i \in \{m, \dots, n\}$.*

For any fixed payments v_n , v_m , and v_0 , an incentive compatible action must satisfy

$$c_i = \gamma_i(v_m - v_0) + \delta_i(v_n - v_0) \quad \text{for all } i \in \{m, \dots, n\}. \quad (1.6)$$

That is, the incentive compatible action is uniquely determined by the measurement rule. Furthermore, if $\gamma_n = \delta_m = 0$ then (1.6) reduces to $c_i = \gamma_i c_m + \delta_i c_n$, which is nondecreasing and is convex, linear, or concave in i over $\{\pi_m, \dots, \pi_n\}$ only if the measurement rule is respectively conservative, unbiased, or aggressive over $\{\pi_m, \dots, \pi_n\}$.

The intuition behind the first result is simple. The agent increases the probability of state π_i until the marginal cost of bringing about that state is equal to the expected marginal payment in that state; that is, $c_i = E[v|\pi_i] - v_0$. Moreover, whereas the contract specifies the degree of potential variation in c_i via v_m and v_n , the measurement rule precisely determines that variation through the selection of (γ_i, δ_i) for each $i \in \{m, \dots, n\}$. Holding the payments fixed with $v_n > v_m$, c_i is increasing in δ_i and decreasing in γ_i . That is, aggressive measurement leads to greater effort, all else equal, because it increases the probability that the agent receives the high payment.⁷

The second statement in Proposition 1.2 is also easily demonstrated. Substituting $\delta_i + \gamma_i = 1$ into (1.6) yields an equation for the unique measurement rule that implements a given action with fixed payments v_n , v_m , and v_0 :

$$\begin{aligned} \delta_i &= \frac{c_i - (v_m - v_0)}{v_n - v_m} & \text{for all } i \in \{m, \dots, n\} \\ \gamma_i &= \frac{(v_n - v_0) - c_i}{v_n - v_m} & \text{for all } i \in \{m, \dots, n\}. \end{aligned} \quad (1.7)$$

⁷This statement should be interpreted with caution, since the principal chooses the payments v_m and v_n with the measurement rule in mind. To see this, notice that if $\delta_i = 1$ for all i then the agent selects an action satisfying $c_i = v_n - v_0$, whereas if $\gamma_i = 1$ for all i then the agent selects an action satisfying $c_i = v_m - v_0$. The equilibrium level of effort is the same under both regimes, since the optimal choice of v_n in the first is equal to the optimal choice of v_m in the second.

If $\delta_m = \gamma_n = 0$, then $\gamma_i = \frac{c_n - c_i}{c_n - c_m}$ is decreasing in c_i and $\delta_i = \frac{c_i - c_m}{c_n - c_m}$ is increasing in c_i . If c_i is increasing convex (concave), then δ_i must be less than (greater than) γ_i on average, which implies that the measurement rule is conservative (aggressive). Figure 1.6 illustrates the result. This correspondence between the convexity of c_i and the direction of bias forms a powerful basis for characterizing optimal measurement rules.

1.3 Stewardship-oriented production, measurement, and contracts

Propositions 1.1 and 1.2 demonstrate that the measurement rule and the contract act as substitutes in determining productive activity, and that the former becomes relatively more important as measurement error increases. I now turn to the characterization of optimal productive actions, measurement rules, and contracts when measurement error is unavoidable, assuming that the objective is to solve the moral hazard problem by maximizing the contracting parties' joint welfare (i.e., the stewardship objective).

I begin by characterizing first and second best actions without measurement error. Both benchmarks can be obtained by analyzing the following program:

$$\begin{aligned}
\max_{p,v} \quad & \sum_{i=0}^N \pi_i p_i - \sum_{i=0}^N h(v_i) p_i \\
\text{s.t.} \quad & \sum_{i=0}^N v_i p_i - c(p) \geq \bar{v} \\
& c_i = v_i - v_0 \quad \text{for all } i.
\end{aligned} \tag{1.8}$$

Substituting the incentive compatibility constraints into the objective function and taking the first order condition with respect to v_0 yields $\lambda = \sum_{i=0}^N h'(v_0 + c_i) p_i$, where λ is the positive Lagrange multiplier on the individual rationality constraint. The first best benchmark can be obtained by considering a risk neutral agent; that is, by setting $h(v) = v$ which yields $h'(\cdot) = \lambda = 1$. The first and second best actions now follow immediately from the first order

condition with respect to p_i :

$$\pi_i - \pi_0 = c_i, \tag{FB_i}$$

$$\pi_i - \pi_0 = (h(v_0 + c_i) - h(v_0)) + (h'(v_0 + c_i) - \lambda)c_{ii}p_i. \tag{SB_i}$$

The left hand side of (SB_{*i*}) represents the marginal productive benefit of increasing p_i . The right hand side represents the marginal cost to the principal, which is the composition of two terms. The first term, $h(v_0 + c_i) - h(v_0)$, represents the change in the distribution of dollar wages resulting from the increase in p_i . The second term, $(h'(v_0 + c_i) - \lambda)c_{ii}p_i$, represents the change in wages needed to ensure that the increase in p_i is incentive compatible while continuing to bind the individual rationality constraint. Since $c_{ii}p_i$ is positive, incentivizing a higher p_i always slackens the individual rationality constraint, thereby allowing the principal to recover at least some of the cost of stronger incentives. In fact, the agent's diminishing marginal utility of wealth allows the principal to recover more than this cost when c_i is small; this can be seen by noting that $\lambda = E[h'(v_0 + c_i)]$, which implies that the second term is negative for c_i roughly below average.

Given Proposition 1.2, I am primarily interested in the concavity or convexity of c_i implied by (FB_{*i*}) and (SB_{*i*}). Since the left hand side of (FB_{*i*}) is increasing linearly in i , the first best action sets c_i increasing linearly in i . In contrast, the first term on the right hand side of (SB_{*i*}) is convex in c_i and thus promotes an action with c_i concave in i . This, given the additive separability of $c(p)$, promotes thinner tails under (SB_{*i*}) than under (FB_{*i*}). The effect of the second term is somewhat ambiguous, since $h'(\cdot)$ and $c_{ii}p_i$ could be concave or convex in c_i , but the term is certainly negative for small c_i and positive for large c_i . Since the right and left hand sides of (SB_{*i*}) must equate, it follows that the second term promotes a larger (smaller) c_i under (SB_{*i*}) than under (FB_{*i*}) when i is small (large). This can be interpreted as a downward shift in

the fundamental earnings distribution. Thus the mean- and variance-reducing effects of moral hazard documented by Holmström (1979), Rogerson (1985), and Sung (1995) are preserved in this nonparametric setting without measurement error. Figure 1.7 illustrates these benchmarks, assuming that the convexity in the first term on the right hand side of (SB_i) dominates any concavity in the second.

Having characterized the first and second best benchmark actions, I am now prepared to analyze optimal actions, contracts, and measurement rules in a second best setting with measurement error. I hereafter refer to this as the third best setting. I substitute (1.7) into the expression for $P(x_i)$ so that the principal jointly chooses an action and its incentive compatible measurement rule in the program given by (1.3). Since δ_i and γ_i must lie within the interval $[0, 1]$, (1.6) requires that $c_i \in [v_{m_j} - v_0, v_{n_j} - v_0]$ for all $i \in \{m_j, \dots, n_j\}$. The modified program can now be written

$$\begin{aligned}
\max_{p, v} \quad & \sum_{i=0}^N \pi_i p_i - \sum_{i=0}^N h(v_i) P(x_i) \\
\text{s.t.} \quad & \sum_{i=0}^N v_i P(x_i) - c(p) \geq \bar{v} \\
& v_{n_j} - v_0 \geq c_i \quad \text{for all } i \in \{m_j, \dots, n_j\} \text{ and } j \in \{1, \dots, J\} \\
& v_{m_j} - v_0 \leq c_i \quad \text{for all } i \in \{m_j, \dots, n_j\} \text{ and } j \in \{1, \dots, J\}.
\end{aligned} \tag{1.9}$$

Let ν_i denote the Lagrange multiplier on the constraint specifying $v_{n_j} - v_0 \geq c_i$ ($\delta_i \leq 1$) and let μ_i denote the multiplier on the constraint specifying $v_{m_j} - v_0 \leq c_i$ ($\gamma_i \leq 1$). I now characterize the solution to this program.

Proposition 1.3. *Assume that $\pi_i - \pi_0$ is linear in i and that $v_n > v_m$ for each $(m, n) \in \{(m_1, n_1), \dots, (m_J, n_J)\}$. Then the principal implements an action and measurement rule sat-*

isfying

$$\pi_i - \pi_0 = (h(v_m)\gamma_i + h(v_n)\delta_i - h(v_0)) + \left(\frac{h(v_n) - h(v_m)}{v_n - v_m} - \lambda \right) c_{ii}p_i + (\nu_i - \mu_i)c_{ii}, \quad (\text{TB}_i)$$

where

$$\sum_{i=m}^n \mu_i = \left(\frac{h(v_n) - h(v_m)}{v_n - v_m} - h'(v_m) \right) P(x_m) \quad \text{and} \quad \sum_{i=m}^n \nu_i = \left(h'(v_n) - \frac{h(v_n) - h(v_m)}{v_n - v_m} \right) P(x_n). \quad (1.10)$$

Moreover, if the agent is risk averse, if $P(x_m) > 0$, if $P(x_n) > 0$, and if c_i is nondecreasing in i , then $c_m = v_m - v_0$ and $c_n = v_n - v_0$.

The right hand side of (TB_i) is composed of three terms, the first two of which have counterparts in (SB_i) . However, it is the third term, $(\nu_i - \mu_i)c_{ii}$, that drives most of the novel results in this chapter.

1.3.1 Earnings asymmetry, all-or-nothing measurement, and bonus caps and floors

Recall that ν_i and μ_i are the Lagrange multipliers on the constraints requiring that $v_m - v_0 \leq c_i \leq v_n - v_0$ for all $i \in \{m, \dots, n\}$. As long as c_i is nondecreasing in i , (1.10) reveals that μ_i increases for small i and ν_i increases for large i as the agent becomes more risk averse or as measurement error raises the probability of extreme earnings. Intuitively, measurement error exposes the agent to additional tail risk, which compels the principal to squeeze v_n and v_m closer together to reduce the risk premium.

Reducing v_n and increasing v_m comes at the cost of underproduction on the high end of Π^j and overproduction on the low end of Π^j . Since c_i is bounded between $v_m - v_0$ and $v_n - v_0$, the third best action exhibits a region just above π_m and just below π_n in which γ_i and δ_i are respectively equal to one, μ_i and ν_i are respectively positive, and the marginal costs are flat. If i is not in either of these regions, then an interior measurement rule is optimal and $\mu_i = \nu_i = 0$,

in which case the third term disappears from (TB_i) .

The resulting action specifies marginal costs that are capped below and above, as illustrated in Figure 1.8. Notice that the total variation in marginal costs under the third best is bounded by the total variation under the second best, a feature that is verified by the following lemma.

Lemma 1.1. *For all $(m, n) \in \{(m_1, n_1), \dots, (m_J, n_J)\}$,*

1. $c_m = v_m - v_0$ *is greater in the third best than in the second best, and*
2. $c_n = v_n - v_0$ *is smaller in the third best than in the second best.*

Given the assumption that $c(p)$ is additively separable in p_i , Lemma 1.1 immediately implies that the lower (upper) tail of the fundamental earnings distribution is thicker (thinner) in the third best than in the second best, which is also illustrated in Figure 1.8.

Corollary 1.1. *If $J = 1$ then the third best action leads to a fundamental earnings distribution with a thinner upper tail and a thicker lower tail than does the distribution implemented by the second best action.*

It also follows immediately that the caps (floors) are implemented by a measurement rule that sets $\delta_i = 1$ ($\gamma_i = 1$), implying a very aggressive (conservative) measurement rule in this region. Figure 1.9 illustrates this feature. Note that whenever the marginal costs fall above (below) the line connecting c_m and c_n , the implementing measurement rule is aggressive (conservative).

Corollary 1.2. *The third best action is implemented by a measurement rule that is conservative over low outcomes and aggressive over high outcomes within each informationally distinct range, a feature of “all-or-nothing” recognition criteria.*

Finally, recall that the incentive compatibility constraint without measurement error specifies $c_i = v_i - v_0$, which hints at the optimality of a contract that also exhibits floors and caps

when $\theta_i \in (0, 1)$. This conjecture can be verified. The following lemmas characterize the optimal contract and measurement rule for a fixed action when $\theta_i \in (0, 1)$; the corollary combines these results with Proposition 1.3 to establish that measurement error creates a demand for caps and floors in the accounting-based bonus plan.

Lemma 1.2. *Suppose that $\theta_i \in (0, 1)$ for all $i \in \{m, \dots, n\}$. Let $\underline{v}_i \equiv E[v|\pi_i, x < x_i]$ denote the expected wage when π_i is realized and measured conservatively, and let $\bar{v}_i \equiv E[v|\pi_i, x > x_i]$ denote the expected wage when π_i is realized and measured aggressively. If the principal chooses v_i and $\delta_i, \gamma_i \in (0, \theta)$ to implement some action p , then he chooses v_i strictly between \underline{v}_i and \bar{v}_i .*

Lemma 1.3. *Suppose that $\theta_i \in (0, 1)$ for all $i \in \{m, \dots, n\}$, with $\delta_m = \gamma_n = 0$. If the principal chooses δ_i, γ_i , and v_i to implement an action in which c_i is capped above and below, then the optimal contract exhibits bonus caps and floors.*

Intuitively, the contractual payment v_i and the aggressive parameter δ_i act as substitutes in incentivizing higher effort. To illustrate, $c_{n-1} = c_n$ can only be incentivized if the agent's expected wage given π_{n-1} is the same as his expected wage given π_n . If $v_{n-1} < v_n$, the only way to accomplish this is to set $\delta_{n-1} = 1$. But $\theta_{n-1} < 1$ precludes such a measurement rule; thus $v_{n-1} = v_n$ is optimal. Similar logic applies to all i satisfying $c_i = c_n$, and a symmetric argument holds for all i satisfying $c_i = c_m$.

Now, the objective function and all constraints in the program characterized by (1.9) are continuous in θ_i , so the optimal action p approaches that characterized by Proposition (1.3) as θ_i increases to one. Since this action specifies marginal costs that are capped below and above, the following corollary now follows immediately from Lemma 1.3.

Corollary 1.3. *The optimal limiting contract as $\theta_i \rightarrow 1$ for all i exhibits bonus caps and floors at the boundaries of each informationally distinct range.*

If $J = 1$, Corollary 1.3 leads to an accounting-based bonus plan that has similar features to that described by Murphy (1999). Specifically, it consists of a single floor, a single incentive zone in which v_i is increasing in i (see Lemma 1.2), and a single cap. The only feature from Murphy (1999) that is missing from my analysis thus far is the hurdle bonus, which is a discrete jump in compensation at the boundary of the bonus floor and the incentive zone. In the next section I show that hurdle bonuses also arise optimally in my setting at predictable thresholds.

1.3.2 Discontinuities, conditional bias, and hurdle bonuses

In the prior section I considered the implications of Proposition 1.3 within informationally distinct ranges. I now consider its across-range implications. Two results immediately follow by simply noting that Lemma 1.1 holds at the boundary of any informationally distinct ranges.

Corollary 1.4. *The third best contract exhibits a hurdle bonus at the boundary of each informationally distinct range relative to the second best contract.*

Corollary 1.5. *The fundamental earnings distribution implemented by the third best action exhibits a discontinuity relative to the distribution implemented by the second best action at the boundary of each informationally distinct range. Moreover, under the economically plausible condition that $c_i > c_{i-1}$ whenever $p_i = p_{i-1}$, there exists a region to the left and to the right of each boundary over which the third best action satisfies p_i declining in i .*

Intuitively, the principal shields the agent from measurement risk by flattening the contract over imprecisely measured regions, which naturally leads to a hurdle bonus at the boundaries of each region. These bonuses incentivize the agent to implement a fundamental earnings distribution exhibiting a discontinuity at the bonus threshold. Figure 1.10 illustrates this result.

Corollary 1.2 demonstrates that third best measurement rules are conservative with respect to low outcomes and aggressive with respect to high outcomes within any informationally

distinct range. I now consider how average bias varies across ranges. To accomplish this, I turn to the first two terms on the right hand side of (TB_i) and consider how they differ from those on the right hand side of (SB_i) . The first term, $h(v_m)\gamma_i + h(v_n)\delta_i - h(v_0)$, is increasing linearly in δ_i and therefore in c_i ; thus the demand for concave c_i promoted by the first term in (SB_i) is neutralized by measurement error. Given Proposition 1.2, this implies that the optimal measurement rule is completely determined by the second and third terms in (TB_i) .

The second term in (TB_i) , $\left(\frac{h(v_n)-h(v_m)}{v_n-v_m} - \lambda\right) c_{ii}p_i$, differs only from the second term in (SB_i) in that the marginal cost of incentivizing p_i is constant rather than increasing in $i \in \{m, \dots, n\}$. This is because additional effort is incentivized by increasing δ_i , leading to a linear increase in expected dollar wages, as opposed to increasing v_i , which leads to a convex increase in expected dollar wages. By the convexity of $h(\cdot)$, the marginal cost of incentivizing p_i for small (large) i is larger (smaller) in the third best than in the second best. This causes probability mass to shift from the lower end of Π^j to the upper end of Π^j for all j .

More importantly, the second term in (TB_i) preserves the property that the net marginal cost of providing incentives is negative for small j and positive for large j , since the cost of increasing incentives over small (large) j can (cannot) be fully recovered by tightening the slack in the individual rationality constraint. This implies that the second term in (TB_i) changes concavity exactly once over Π , provided the concavity of $c_{ii}p_i$ is robust to p . This observation motivates the following corollary.

Proposition 1.4. *Suppose that for any p , $c_{ii}p_i$ is concave (convex) in i . If ν_i and μ_i are sufficiently close to zero for all i , then there exists some j^* such that the third best action satisfies c_i concave (convex) over Π^j for all $j < j^*$ and convex (concave) over Π^j for all $j \geq j^*$. That is, given Proposition 1.2, the direction of bias optimally reverses at a single threshold. For example,*

1. *If $c_{ii}p_i$ is strictly concave in i and if $\pi_{m_{j^*}} = 0$ then the realization rule is optimal.*

2. If $c_{ii}p_i$ is strictly concave in i over gains and is linear (weakly concave) in i or c_i over losses then conditional conservatism is optimal.
3. If $c_{ii}p_i$ is linear in i or c_i then fair value is optimal.

Of course, the usefulness of Proposition 1.4 depends on the existence and descriptiveness of cost functions in which the properties of $c_{ii}p_i$ are relatively robust to p . The following example demonstrates that such cost functions exist.

Example 1.1. Suppose $c(p) = \sum_{i=1}^N b_i p_i^{1+a_i}$ where $b_i > 0$ for all i and $a_i > 0$ is nonincreasing in i . Then $c_{ii}p_i = a_i c_i$, and it is straightforward to show the following:

1. For any p , if a_i is sufficiently convex (concave) then $c_{ii}p_i$ is also convex (concave).
2. If a_i is constant in i , then $c_{ii}p_i$ is linear in c_i and fair value is optimal.
3. If a_i is weakly concave in i , then conservative measurement of high outcomes is optimal.

The first observation is that the type of robust cost functions needed to invoke Proposition 1.4 exist and have economically defensible properties. The second observation implies that the popular quadratic cost function creates a stewardship demand for unbiased measurement. The third observation is that conservative measurement of high outcomes is optimal whenever a_i is concave or linear in i , the latter being an arguably neutral ex ante assumption.

A heuristic proof of these three observations follows. Define $a(i) \equiv a_i$, $c(i) \equiv c_i$, and $C(i) \equiv a(i)c(i)$. Assume for the moment that i is chosen from a continuum and that $a(\cdot)$ and $c(\cdot)$ are twice continuously differentiable functions. Then

$$C''(i) = a(i)c''(i) + 2a'(i)c'(i) + a''(i)c(i), \quad (1.11)$$

where $a(i) > 0$, $c(i) > 0$, $a'(i) \leq 0$, and $c'(i) > 0$. If $a''(i)$ is sufficiently positive (negative), then $C''(i)$ is also positive (negative), proving the first observation. The second observation

follows directly from Proposition 1.4. Finally, suppose to the contrary that $a(i)$ is decreasing and weakly concave and that the optimal measurement rule does not treat high outcomes conservatively; that is, $a'(i) < 0$, $a''(i) \leq 0$, and $c''(i) \leq 0$. Then each of the terms on the right hand side of (1.11) is nonpositive and the second is strictly negative, so $C''(i) < 0$. But $c''(i) \leq 0$ and $C''(i) < 0$ over high outcomes contradicts Proposition 1.4, thereby proving the result.

The impact of the third term in (TB_i) on average bias over Π^j can be determined by (1.10); notice that the magnitude of the flat regions just above and below π_m and π_n are loosely proportional to $P(x_m) = \sum_{i=m}^n p_i \gamma_i$ and $P(x_n) = \sum_{i=m}^n p_i \delta_i$, respectively. If p is uniform, then the flat region above π_m is larger (smaller) than the flat region below π_n whenever c_i is increasing convex (concave) between these regions. In this case, the third term in (TB_i) enhances the demand for the measurement rule implied by Proposition 1.4.

On the other hand, if p_i is increasing (decreasing) in i over Π^j , then $P(x_m)$ is relatively smaller (larger) than $P(x_n)$, all else equal. In turn, (1.10) yields a flat region just above π_m that is smaller (larger) than the region below π_n , which implies a demand for aggressive (conservative) reporting. It follows that if p is bell-shaped and centered at zero, so that p_i is increasing over losses and decreasing over gains, then the third term in (TB_i) promotes the realization rule.

Conclusion

I have demonstrated that a very general version of the moral hazard problem is able to rationalize empirically descriptive earnings distributions, measurement rules, and contracts. First, I follow Holmstrom and Milgrom (1987) and grant the agent intricate control over the production function by allowing him to directly choose the probability of each outcome. Second, I introduce a novel analytical representation of an accounting system that admits a large class of

measurement rules applied in practice. In this setting, I show that the measurement rule and the contract act as substitutes in inducing productive activity.

My first set of results relies on a form of measurement error that increases the dispersion of reported earnings. I show that this type of measurement error promotes fundamental earnings distributions with asymmetric tails, measurement rules with “all-or-nothing” features, and contracts with bonus caps and floors. My second set of results relies on the ability of the accounting system to cleanly distinguish fundamental performance that exceeds some threshold from performance that falls short. In this setting, I show that measurement error promotes hurdle bonuses in the optimal contract, discontinuities in the fundamental earnings distribution, and measurement rules that are conditionally biased.

Thus far, I have assumed that the principal has control over the contract and the measurement rule, and is thus able to easily induce any productive action that he chooses. In the next chapter, I introduce an accounting regulator who is able to exert influence over the measurement rule. In this setting, the principal must design the contract to mitigate the shortcomings of a suboptimal measurement rule, but he only partially does so in equilibrium. Thus the accounting regulator is able to influence the actions agreed upon by the contracting parties, and therefore has some control over aggregate production in the economy.

Chapter 2

What You Measure is What You Get:

Standardizing Production through Uniform Reporting

The merits of uniform accounting standards have been hotly debated in recent decades. While comparability – which specifies that like things look alike and different things look different – is widely acknowledged to be a worthwhile aim, academics and standard setters have pointed out that uniformity does not necessarily imply comparability. On the one hand, uniformity could keep firms from accounting for similar transactions differently, while on the other hand, uniformity could force dissimilar phenomena to look alike (FASB 2010). While these arguments are important, both assume that transaction characteristics are independent of the accounting standards in place. Indeed, the accountant’s favorite proverb “what you measure is what you get” suggests an alternative role for uniform reporting: it may not just make unlike things *look* more alike, but it may actually cause them to *be* more alike in the first place.

In this chapter I present a case in which uniform reporting causes firms with different risk preferences to adopt similar risk profiles. Whereas in Chapter 1 I assume that productive activity, measurement rules, and contracts are all chosen with the purpose of solving the moral hazard problem in the context of a single principal-agent relationship, this chapter allows for multiple heterogeneous principal-agent pairs who contract on a measurement rule that is mandated by an accounting regulator. I assume that firms are inclined to contract on publically available financial reports because (i) information production and verification is costly, (ii)

contracts that reward agents following poor public performance are unlikely to be sustainable, and (iii) complete separation of financial and managerial accounting systems is unlikely to be optimal and is not empirically descriptive (e.g., Hemmer and Labro 2008, Dichev et al. 2013).

The first modification to my set-up from Chapter 1 is the introduction of heterogeneous firms. To fix ideas, I consider firms whose preferences vary over downside risk. Extremely poor fundamental performance in any given period is likely to lead to differential long term consequences depending on the firm's financial health and competitive environment. For example, some firms may suffer from liquidity problems in the event of poor short-term performance, which would force the firm to tap into external financing sources at a relatively higher cost. Moreover, temporary poor performance could keep the firm from gaining a strong competitive foothold in certain markets. In contrast, limited liability or the liquidation option could dampen the consequences of poor short-term performance to the shareholders of particular firms. Based on these arguments, aversion to downside risk is likely to depend on factors that vary across firms.

The second modification to my set-up is the introduction of an accounting regulator. A foundational insight from Chapter 1 is that measurement error renders the agent's action less sensitive to variation in the contract and more sensitive to the properties of the measurement rule. In this chapter, the contract is designed by the principal, whereas the measurement rule is mandated by the accounting regulator. Thus the regulator's influence over production is tied to the degree of measurement error in the accounting system.

In a multiperiod version of the model, I derive an explicit link between timely measurement and aggregate measurement error based on the assumption that uncertainty is resolved over time. Thus an accounting regulator can plausibly choose his sphere of influence over productive activity by mandating the timeliness of financial reporting. Importantly, this also implies that my analytical representation of an accounting system is able to capture features of historical

cost, current value, timely loss recognition, and timely gain recognition through appropriate specifications of the noise parameters.

I confine my attention to the effect of a single pervasive measurement rule, timely loss recognition, on productive activity. The model predicts that a uniform application of timely loss recognition reduces (increases) downside risk within the subset of firms that are less (more) averse to it. That is, timely loss recognition reduces variation in downside risk across firms. While prior empirical studies have demonstrated that timely loss recognition promotes the selection of more profitable projects on average (e.g., Bushman, Piotroski, and Smith 2011), this chapter provides cross-sectional predictions pertaining to the same tests. For example, I predict that timely loss recognition actually increases the selection of bad projects among firms with poor liquidity.

2.1 Equating timeliness with measurement error in a multiperiod model

Consider a T period version of the model described in Chapter 1, where the managed asset has a fixed operating cycle of two periods. While the asset's period t change in fundamental value $\pi^t \in \{\pi_0, \pi_1, \dots, \pi_N\}$ is not immediately observable, the accounting system produces a timely but noisy estimate $x^t \in \{x_0, x_1, \dots, x_N\}$. For full generality, I impose no restrictions on the conditional probabilities $\Pr(x^t|\pi^t)$, so (1.1) and (1.2) are allowed but not required. In period $t+1$, the results from period t production are fully realized, and any measurement error generated by the accounting system in period t is completely reversed. Thus period t earnings can be written

$$y^t \equiv \begin{cases} x^1 & \text{if } t = 1 \\ x^t + (\pi^{t-1} - x^{t-1}) & \text{if } t \in \{2, \dots, T\} \\ \pi^T - x^T & \text{if } t = T + 1. \end{cases} \quad (2.1)$$

That is, earnings consist of a timely but noisy component x^t generated during the current period and a reversing accrual $\pi^{t-1} - x^{t-1}$ that corrects measurement error from the prior period. Notice that this earnings process guarantees that $\sum_{t=1}^{T+1} y^t = \sum_{t=1}^T \pi^t$, thereby capturing the property that accrual accounting simply allocates economic income across periods.

While total economic income can be deduced from total earnings, a perfect deduction of $\{\pi^t\}_{t=1}^T$ is not generally possible. To illustrate, if $x^t = \pi^t + \epsilon^t$ approximates the current period's economic income with some noise, then in any given period $y^t = \pi^t + \epsilon^t - \epsilon^{t-1}$ is a noisy measure of current period income confounded by reversals from prior periods. If aggregation precludes a distinction between measurement error in the current period and error reversals from the prior period, then a perfect deduction of $\{\pi^t\}_{t=1}^T$ is impossible.

While a timely, unbiased measure of economic income introduces noise to the earnings process that cannot be fully filtered out over time, an untimely measure allows for a perfect deduction of $\{\pi^t\}_{t=1}^T$. To see this, note that under historical cost accounting with $x^t = 0$ for all t , (2.1) reduces to $y^t = \pi^{t-1}$.¹ It follows immediately from Holmström's (1979) sufficient statistic condition that historical cost accounting is at least weakly preferred by the principal to any other measurement rule in this setting since it completely eliminates measurement error. The above discussion constitutes a proof of the following result.

Proposition 2.1. *Suppose that realized income is a sufficient statistic for realized and unrealized income with respect to the agent's action, that uncertainty is resolved deterministically over a fixed operating cycle, and that current period measurement error is aggregated with accrual reversals from prior periods. Then historical cost accounting eliminates measurement error and is optimal from a stewardship perspective.*

Deterministic accrual reversals over multiple periods creates a moving support problem

¹If the accounting system adheres to (1.1) and (1.2), then this can be done by applying Definition 1.2 with $J = 2$, $\pi_{n_1} \lesssim \pi_0 = 0 \lesssim \pi_{m_2}$, and $\theta_i = 1$ for all i .

that complicates the characterization of optimal actions and contracts. While such a characterization is unnecessary for Proposition 2.1, I propose a reduced form representation that allows timeliness to be captured in a static model. Specifically, I rely on the link between timeliness and measurement error derived in the multiperiod model to exogenously equate timeliness with measurement error in the static model from Chapter 1.

Definition 2.1. *Consider the static model described in Chapter 1 with an accounting system satisfying (1.1) and (1.2), and assume that fundamental gains and losses are informationally distinct. Then,*

1. *An accounting system applies current value if it is noisy over gains and losses.*
2. *An accounting system applies historical cost if it is precise over gains and losses.*
3. *An accounting system applies timely loss recognition if it is precise over gains and noisy over losses.*

2.2 Influencing downside risk through timely loss recognition

With Definition 2.1 in hand, I now return to the more tractable static model from Chapter 1. Consider K principal-agent pairs, one for each firm, and suppose that each principal has nonlinear preferences over short-term performance. As previously discussed, poor short-term performance can lead to liquidity issues, trigger the liquidation option, or encourage asset substitution due to limited liability, thereby inducing nonlinear preferences.

I model principal k 's nonlinear preferences through an additive component equal to ϕ_i^k that accrues to the k^{th} principal when π_i is realized. Thus the principal's utility is equal to $\pi_i + \phi_i^k - E[h(v)|\pi_i]$ when the current period's fundamental performance is equal to π_i . Now, a firm that is very averse to downside risk can be modeled by choosing ϕ_i^k such that $\pi_i + \phi_i^k$ becomes very small as i declines, whereas a firm that is less averse to downside risk can be

modeled by choosing ϕ_i^k such that $\pi_i + \phi_i^k$ “levels off” over small i . A parsimonious way to capture this is to assume that $\pi_i + \phi_i^k$ is increasing concave for firms that are particularly averse to downside risk and increasing convex for firms that are not.

I also assume that the agent is risk neutral. This assumption removes risk-sharing considerations from the optimal contract and measurement rule, allowing me to focus exclusively on the demand for different risk profiles based on the nonlinear component ϕ_i^k . In this case, (TB_{*i*}) reduces to

$$\phi_i^k + \pi_i - \pi_0 = c_i. \quad (2.2)$$

That is, if the principal could choose the measurement rule, he would choose it such that the convexity of c_i matches the convexity of ϕ_i^k .

Of course, the measurement rule is chosen by the accounting regulator, not by the principal. The following proposition presupposes a regulator who, for whatever reason, desires to mandate timely loss recognition.

Proposition 2.2. *Suppose $J = 2$ with $\pi_{m_2} = 0$, and suppose that the accounting regulator mandates timely loss recognition with δ_i and γ_i chosen to satisfy $E[x|\pi_i \in \Pi^1] = \pi_i$. Then the k^{th} principal chooses contractual payments v_{m_1} and v_{n_1} to satisfy the following equation:*

$$0 = \sum_{i=m_1}^{n_1-1} (\phi_i^k + \pi_i - \pi_0 - c_i) \frac{\gamma_i}{c_{ii}} \quad \text{and} \quad 0 = \sum_{i=m_1+1}^{n_1} (\phi_i^k + \pi_i - \pi_0 - c_i) \frac{\delta_i}{c_{ii}}, \quad (2.3)$$

where $c_i = \gamma_i v_{m_1} + \delta_i v_{n_1} - v_0$ by (1.6). Moreover, if $\phi_i^k + \pi_i$ is increasing concave (convex) in i , then there exists some $l \geq 1$ such that c_1, c_2, \dots, c_l are chosen strictly greater (smaller) under (2.3) than under (2.2).

Notice that if (2.2) were satisfied for firm k , then each term in (2.3) would be equal to zero. However, since the principal does not choose the measurement rule, (2.2) will not in general

be satisfied for all i . Therefore, the best the principal can do is choose v_{m_1} and v_{n_1} such that (2.2) is satisfied *on average*. Intuitively, the principal chooses the boundary payments to “fit” the implemented marginal costs, like a regression line, to the costs he would have chosen if he’d had control over the measurement rule. This “fitting” causes the lower tail of marginal costs to increase (decrease) for firms that are more (less) averse to downside risk. In other words, a uniform application of timely loss recognition reduces variation in downside risk across firms. Figure 2.1 illustrates this result.

Conclusion

In this chapter I have demonstrated two results. First, if uncertainty is resolved over time then an accounting regulator can influence productive activity in the economy by mandating timely measurement. Second, I have shown that timely loss recognition reduces the variation in downside risk across firms: specifically, firms that are less averse to downside risk contract for less risky actions, whereas firms that are more averse to downside risk are induced to contract for riskier actions.

While perhaps significant in its own right, this chapter is meant to be more illustrative than prescriptive. Specifically, my objective here is to show that this framework offers opportunities to better understand the effect of financial reporting regulation on economic activity. With further development, it is my hope that this framework will offer additional insight into issues that have long interested accounting researchers such as uniformity, comparability, and standard setting.

Chapter 3

An Optimistic Case for Pessimistic Measurement: Disabling Moral Hazard through Unconditional Conservatism

Criticisms of conservative reporting tend to be levied against its unconditional, rather than its conditional, form. While its undesirability from a valuation perspective is perhaps self-evident, even proponents of the stewardship or debt contracting objectives tend to favor conditional conservatism over its unconditional counterpart. The following argument made by Ball and Shivakumar (2005) is illustrative:

[U]nconditional biases reduce opportunities to account in a conditionally conservative fashion (for example, writing off assets at acquisition eliminates the opportunity to impair them at the time of economic losses). Contracting based demand for a known unconditional bias thus seems unlikely. Further, an unconditional bias of unknown magnitude introduces randomness in decisions based on financial information and can only reduce contracting efficiency.

Such statements call into question the longstanding practice of leaving most intangible assets, such as research and development (R&D), off the balance sheet. Such assets are not immaterial, since growth and long-term profitability often depend critically on their successful management. If GAAP is efficient, then these assets must possess special characteristics that imply the optimality of unconditionally conservative measurement.

In this chapter I provide sufficient conditions for conservative bias and noise to be strictly desirable given the stewardship objective. Specifically, if an optimistic agent has a rich set of value-augmenting actions at his disposal, then a penalty contract designed after Mirrlees (1974), which is characterized by a flat wage less a penalty in the event of very low earnings, can induce first best allocations only if the measurement system is extremely noisy and conservatively biased.

The result relies on two main ingredients. The first is an assumption first employed by Holmstrom and Milgrom (1987), namely that the agent has a large degree of flexibility when choosing which fundamental earnings distribution to implement. Such flexibility causes a penalty contract to lead to particularly undesirable actions when measurement is precise. This occurs because the agent can avoid penalties in many different ways, and the cheapest penalty-avoidance strategy is the one that allows for maximum shirking. Intuitively, the agent works until fundamental earnings lie “just above” the penalty zone, then withholds productive effort thereafter as argued by Holthausen, Larcker, and Sloan (1995).

While abandoning the flat wage scheme overcomes this problem, it is not the only recourse. Productive activity can be recovered by substituting variation in the contract with variation in the measurement system. Noise disrupts the agent’s otherwise intricate control over the reported earnings distribution, and conservatively-biased noise puts any agent in danger of incurring the penalty, no matter how precise his control over fundamentals. To the extent that closing entries governed by conservative measurement rules deflate reported earnings by some seemingly random amount, agents are inclined to exert productive effort even over ranges in which compensation is flat.

While noise and conservative bias are able to render the first best action incentive compatible via a penalty contract, this does not by itself imply that such a contract is Pareto efficient: a penalty contract may impose too much risk on the agent. Thus the second ingredient in my

model is that the agent be optimistic about the odds of avoiding a penalty-triggering earnings report. This optimism causes the agent to charge a smaller risk premium when offered a penalty contract, thereby ensuring the joint optimality of the first best action and the penalty scheme as the probability of incurring the penalty goes to zero.

There is a burgeoning literature on CEO overconfidence and optimism. According to a count by Malmendier and Tate (2015), “about two-dozen articles in top economics and finance journals have been published on the topic” since 2005, and I have identified a handful of related papers in top accounting journals over the past five years. Prior analytical studies in these literatures operationalize CEO optimism (overconfidence) by introducing upward bias to the perceived mean (precision) of a noisy signal. Given the nonparametric nature of the production function in my setting, I take a more flexible approach. I assume that the agent is a rank dependent expected utility (RDEU) maximizer who attaches decision weights to each objective probability. Characterizations of optimism and pessimism are standard in the RDEU literature. Loosely speaking, an optimistic agent tends to overweight the probability of good outcomes and underweight the probability of bad outcomes.

According to Machina (1994), rank dependent expected utility is “the most natural and useful modification of the classical expected utility formula,” and Starmer (2000) observes that it has “proved to be one of the most popular among economists.” Like expected utility theory, RDEU has an axiomatic development with many pleasing normative properties.¹ For example, preferences under RDEU are transitive and they respect first order stochastic dominance. Another appealing feature is that expected utility theory is a special case of RDEU, so it is straightforward in many situations to evaluate whether classical results extend to the more general framework.

¹The critical feature of RDEU is a weakening of the independence axiom, which is without a doubt the most often questioned and violated axiom in expected utility theory.

In sum, I show that managerial optimism enables the reduction and, in limiting cases, the elimination of agency costs when paired with unconditionally conservative measurement rules. Going beyond the notion that conservatism “offsets managerial biases” (see, for example, Watts 2003), my model provides the additional insight that optimism and unconditional conservatism interact to create simple and efficient contracts.

This finding leads to two broad empirical predictions. First, optimistic agents receive less variable compensation and demand lower risk premia than do neutral or pessimistic agents. This is consistent with Otto (2014), who finds that optimistic CEOs receive smaller stock option grants, fewer bonus payments, and less total compensation than do their peers. Second, assets are more likely to be measured in an unconditionally conservative fashion (i) if their values can be augmented through many different types of productive activities and (ii) if their stewards are optimistic with respect to reported earnings. Hirshleifer, Low, and Teoh (2012) find that firms with optimistic CEOs invest more in innovation and achieve greater innovative success for given research and development expenditures. To the extent that these optimistic, innovative agents have a rich set of value-augmenting actions at their disposal, the model rationalizes unconditionally conservative treatment of research and development and other innovation-dependent intangibles.

3.1 A rank dependent expected utility representation of optimism

The model in this chapter is very similar to that employed in Chapter 1, so in the interest of parsimony I outline only the main differences here. The most notable addition to the basic setup is a very low earnings outcome denoted $\underline{x} < x_0 < x_1 < \dots$, which can be interpreted as the average of all possible earnings realizations below x_0 . The accounting system is as described before, except that $\pi_0 \in \Pi^1$, $J = 1$ (so j no longer indexes informationally distinct ranges), and

$\delta_i = 0$ for all i ; that is, only conservative biases are permitted.² Define the set $\Gamma \equiv \{\gamma_0, \dots, \gamma_N\}$. Now, if π_i is realized then the accounting system measures π_i *conservatively* with probability γ_i and *neutrally* or *without bias* with probability $1 - \gamma_i$. Again, if π_i is measured conservatively, then x is equal to the largest $\pi < \pi_i$ that is measured neutrally. Figure 3.1 illustrates this altered structure.

Denote $P(x_i) \equiv \Pr(x_i|p, \Gamma)$. With the above alterations, (1.1) and (1.2) reduce to

$$P(x_i) = (1 - \gamma_i) \sum_{k=i}^N p_k \prod_{j=i+1}^k \gamma_j \quad \text{and} \quad P(\underline{x}) = \sum_{k=0}^N p_k \prod_{j=0}^k \gamma_j, \quad (3.1)$$

using the convention that $\prod_i^k(\cdot) = 1$ if $i > k$. It is straightforward to verify that

$$\begin{aligned} \frac{\partial P(\underline{x})}{\partial p_k} &= \Pr(\underline{x}|\pi_k) - \gamma_0 = -\gamma_0 \left(1 - \prod_{j=1}^k \gamma_j\right), \\ \frac{\partial P(x_0)}{\partial p_k} &= \Pr(\underline{x}|\pi_k) - (1 - \gamma_0) = -(1 - \gamma_0) \left(1 - \prod_{j=1}^k \gamma_j\right), \end{aligned} \quad (3.2)$$

$$\frac{\partial P(x_i)}{\partial p_k} = \Pr(x_i|\pi_k) = \begin{cases} 0 & \text{if } k < i \\ (1 - \gamma_i) & \text{if } k = i \\ (1 - \gamma_i) \prod_{j=i+1}^k \gamma_j & \text{if } k > i > 0. \end{cases}$$

Mirrlees (1974) obtains approximate first best allocations through penalty contracts by sending the penalty zone to negative infinity while increasing the size of the penalty. A symmetric approach for bonuses is ineffective if the agent exhibits decreasing marginal utility of wealth. To replicate Mirrlees' strategy, I need to take the lower tail of the reported distribution

²Aggressive biases are benign in this chapter and lead to no Pareto improvements, so I omit this feature for the sake of simplicity.

to negative infinity, which I accomplish parsimoniously by taking $\underline{x} \rightarrow -\infty$. The penalty is administered upon the occurrence of report \underline{x} .

I now introduce a *rank dependent expected utility (RDEU)* maximizing agent. My brief summary of RDEU draws heavily on the excellent exposition by Quiggin (1992). RDEU is characterized by a transformation of probabilities into decision weights. The weights are determined by a cumulative weighting function $q : [0, 1] \rightarrow [0, 1]$ that takes into account both the magnitude of each probability and the ranking of each outcome. Whereas an expected utility maximizing agent solves

$$\underline{v} P(\underline{x}) + \sum_{i=0}^N v_i P(x_i) - c(p),$$

an RDEU maximizing agent solves

$$\underline{v} g(P) + \sum_{i=0}^N v_i g_i(P) - c(p),$$

where $P \equiv (P(\underline{x}), P(x_0), \dots, P(x_N))$, the cumulative mass function is denoted $\rho_i \equiv \Pr(x \leq x_i)$ and $\underline{\rho} \equiv \Pr(x \leq \underline{x}) = P(\underline{x})$, and

$$\begin{aligned} g_i(P) &\equiv q(\rho_i) - q(\rho_{i-1}) \text{ for all } i > 0, \\ g_0(P) &\equiv q(\rho_0) - q(\underline{\rho}), \text{ and} \\ \underline{g}(P) &\equiv q(\underline{\rho}). \end{aligned} \tag{3.3}$$

The weights assigned to each probability can be assessed by considering the slope of q over each

segment of the cumulative mass function. (3.3) implies that

$$P(x_i) = \rho_i - \rho_{i-1} \quad \left\{ \begin{array}{ll} > g_i(P) & \text{if } \frac{q(\rho_i) - q(\rho_{i-1})}{\rho_i - \rho_{i-1}} < 1 \\ = g_i(P) & \text{if } \frac{q(\rho_i) - q(\rho_{i-1})}{\rho_i - \rho_{i-1}} = 1 \\ < g_i(P) & \text{if } \frac{q(\rho_i) - q(\rho_{i-1})}{\rho_i - \rho_{i-1}} > 1. \end{array} \right.$$

In words, the agent underweights segments of the cumulative mass function over which $q'(\rho) < 1$ and overweights segments over which $q'(\rho) > 1$. It follows that an agent with a convex weighting function increasingly underweights less desirable outcomes and increasingly overweights more desirable outcomes. That is, he is optimistic. Conversely, an agent with a concave weighting function increasingly overweights less desirable outcomes and increasingly underweights more desirable outcomes. That is, he is pessimistic. Finally, an expected utility maximizer has a linear weighting function. These cases are illustrated in Figure 3.2.

The functional form associated with RDEU was independently discovered by several different researchers, the first being Quiggin (1982). He provided the first axiomatic basis for the theory and demonstrated that the decision weights $g_i(P)$ must depend on the cumulative rather than individual probabilities if violations of stochastic dominance are to be avoided.

One rediscovery is worth mentioning. Schmeidler (1989) and Gilboa (1987) develop a generalized expected utility theory often referred to as *Choquet Expected Utility*, named after its use of Choquet integrals in calculating expectations with non-additive subjective probabilities. Such subjective probabilities tend to self-manifest in problems characterized by *ambiguity*, or uncertainty regarding the distribution of outcomes. This generalization was later shown to be equivalent to RDEU when preferences respect first order stochastic dominance (see Wakker 1990).

Segal (1987, 1990) argues that ambiguity can be modeled as a two-stage lottery, where the first lottery determines the probability distribution over outcomes and the second lottery

produces an outcome from that distribution. Segal derives some fairly restrictive sufficient conditions for a single stage lottery to be preferred to a two stage lottery, and one of these conditions is pessimism. When these conditions are not satisfied, an agent may actually prefer an ambiguous lottery to an unambiguous one.

My model shares some common features with Segal's. Segal and I both assume that the agent is a rank dependent expected utility maximizer. Reported earnings in my model is, in fact, the result of a two-stage lottery: the first lottery produces a fundamental outcome which determines the probability distribution over earnings reports, and the second lottery produces an earnings report from that distribution. While I do not formalize the connection here, it seems plausible that my model can be reinterpreted as one of earnings ambiguity under Segal's framework. If so, a manager who is optimistic may prefer to obtain only a high-level understanding of the accounting process (that is, he prefers ambiguity) in order to avoid information that is inconsistent with his optimistic decision weights.

3.2 Penalty contracts with negligible risk premia

I am now prepared to investigate the impact of noise, conservative bias, and optimism on obtainable allocations in a moral hazard setting. The principal's program is given by

$$\begin{aligned}
 \max_{p, v_i} \quad & \sum_{i=0}^N \pi_i p_i - \sum_{i=0}^N h(v_i) P(x_i) - h(\underline{v}) P(\underline{x}) \\
 \text{s.t.} \quad & \sum_{i=0}^N v_i g_i(P) + \underline{v} g(P) - c(p) \geq \bar{v} \\
 & p \in \operatorname{argmax}_{\tilde{p}} \sum_{i=0}^N v_i g_i(\tilde{P}) + \underline{v} g(\tilde{P}) - c(\tilde{p}).
 \end{aligned} \tag{3.4}$$

Notice from (3.1) and (3.3) that $P(x_i)$ is a linear function of p_k whereas $g_i(P)$ is a nonlinear function of P . Absent additional assumptions, the incentive compatibility constraint's first order conditions are necessary but not sufficient for an interior solution; second order conditions must

be verified as well. It turns out that the second order conditions are automatically satisfied under a penalty contract if the agent is optimistic.

Proposition 3.1. *Suppose the principal offers a penalty contract satisfying $v_i = v > \underline{v}$ for all i . If $q''(\underline{\rho}) \geq 0$, then the agent's first order conditions are necessary and sufficient for an interior solution to the incentive compatibility constraint. Moreover, the p_k^{th} first order condition specifies*

$$1 - \prod_{j=1}^k \gamma_j = \frac{c_k}{q'(P(\underline{x}))(v - \underline{v})\gamma_0}. \quad (3.5)$$

Notice from (3.5) that if $\gamma_i = 0$ for some i , then $c_i = c_{i+1} = \dots = c_N = q'(P(\underline{x}))(v - \underline{v})\gamma_0$. This is because any outcome above π_i guarantees that the penalty will not obtain, and the cheapest way to ensure penalty avoidance is to equate the marginal costs in this region. Such actions are very undesirable: they are the types of actions chosen by the agent absent incentives.³ It follows that any accounting system that renders a penalty contract incentive compatible while promoting productive activity *must* allow for conservative measurement.

In contrast, a conservative system is capable of inducing $c_1 < c_2 < \dots < c_N$ because it leaves the manager exposed to the penalty no matter which fundamental outcome is realized. The manager has a natural incentive to allocate more effort towards outcomes that are far from \underline{x} because these outcomes are less likely to result in a penalty-triggering earnings report. The desired variation in c_i can be induced by a penalty contract if the degree of conservatism is correctly calibrated for each outcome. While many incentive compatible calibrations of the measurement system and penalty are possible, for illustrative purposes I focus on the system that is least biased.

³To see this, notice that if $\gamma_0 = 0$ or $v - \underline{v} = 0$ so that a penalty is never inflicted, the marginal costs are all equated to zero.

Proposition 3.2. *Let q be convex and fix $\gamma_N = 0$. If p can be implemented via a penalty contract, then the unique incentive compatible penalty and measurement system satisfy*

$$\begin{aligned} v - \underline{v} &= \frac{c_N}{q'(P(\underline{x}))\gamma_0} \quad \text{and} \\ \gamma_i &= \frac{c_N - c_i}{c_N - c_{i-1}} \quad \text{for all } i > 0. \end{aligned} \tag{3.6}$$

While any conservative accounting system implements *some* action via a penalty contract, this hardly guarantees that the principal will choose to implement that action and offer that contract. There is at least one exception. If a particular accounting system, penalty contract, and weighting function allow the first best action to be implemented with a negligible risk premium, then the principal would be remiss to implement any other action.

Recall from (FB_{*i*}) that the first best action satisfies

$$\pi_k - \pi_0 = c_k \quad \text{for all } k \in \{1, \dots, N\}. \tag{3.7}$$

If one insists on implementing the first best action in a second best world through a penalty contract, then the implementing accounting system from (3.6) is therefore given by $\gamma_i = \frac{\pi_N - \pi_i}{\pi_N - \pi_{i-1}}$ for all $i > 0$. It remains only to characterize γ_0 . Notice that the structure imposed on the conditional probabilities in this framework specifies that $\Pr(x_i | \pi_j)$ is nonincreasing in the distance between i and j . In this spirit, I assume that γ_0 , which is equal to the conditional probability of \underline{x} given π_0 , declines to zero as $\underline{x} \rightarrow -\infty$. While any such relationship between γ_0 and \underline{x} is sufficient for the following results, in the interest of consistency I assume that $\gamma_0 \equiv \frac{\pi_N - \pi_0}{\pi_N - \underline{x}}$ takes the same functional form as the other conservative parameters. This assumption implies that $\Pr(\underline{x} | \pi_k) = \prod_{i=0}^k \gamma_i = \frac{\pi_N - \pi_i}{\pi_N - \underline{x}}$ for all k . With this accounting system in place, it follows from

Proposition 3.2 that

$$P(\underline{x}) = \frac{\sum_{k=0}^N p_k(\pi_N - \pi_k)}{\pi_N - \underline{x}} \quad \text{and} \quad v - \underline{v} = \frac{\sum_{k=0}^N p_k(\pi_N - \pi_k)}{q'(P(\underline{x}))P(\underline{x})}. \quad (3.8)$$

Since $\lim_{\rho \rightarrow 0} q'(\rho)$ is bounded between zero and one if the agent is optimistic, it is immediate that $P(\underline{x}) \rightarrow 0$ and $v - \underline{v} \rightarrow \infty$ as $\underline{x} \rightarrow -\infty$. It remains only to assess limiting behavior of the risk premium.

Proposition 3.3. *Suppose that the following conditions are satisfied:*

1. $q''(\rho) \geq 0$ for all $\rho \in [0, 1]$.
2. $\gamma_0 = \frac{\pi_N - \pi_0}{\pi_N - \underline{x}}$.
3. $\gamma_i = \frac{\pi_N - \pi_i}{\pi_N - \pi_{i-1}}$ for all $i > 1$.

Then the risk premium is proportional to $\frac{q(P(\underline{x}))}{q'(P(\underline{x}))P(\underline{x})}$.

If the agent is an expected utility maximizer, then $\frac{q(P(\underline{x}))}{q'(P(\underline{x}))P(\underline{x})} = 1$ for all \underline{x} . It follows that the desirability of a penalty contract can be assessed by comparing this expression with unity when the agent is optimistic versus pessimistic. Notice that

$$\frac{q(\rho)}{q'(\rho)\rho} < 1 \quad \Longleftrightarrow \quad \frac{q(\rho) - q(0)}{\rho - 0} < q'(\rho),$$

which is equivalent to the convexity of q . Thus the risk premium charged by an optimistic agent is less than the risk premium charged by an expected utility maximizing or pessimistic agent. The following examples illustrate that the risk premium can be made arbitrarily small by making the agent sufficiently optimistic or by taking the penalty zone to negative infinity.

Example 3.1. *Let $q(\rho) = \rho^\alpha$. Then the risk premium is proportional to $\frac{q(\rho)}{q'(\rho)\rho} = \frac{1}{\alpha} \xrightarrow{\alpha \rightarrow \infty} 0$.*

Since the risk premium can be made arbitrarily small by taking $\alpha \rightarrow \infty$, a penalty contract with the specified unconditionally conservative accounting system dominates any other contract

and measurement rule in the limit. Moreover, the manager becomes more and more optimistic that π_N will materialize ($g_N(P) \rightarrow 1$) as $\alpha \rightarrow \infty$.

For the weighting function in Example 3.1, the risk premium only shrinks by making the agent increasingly optimistic. The next example demonstrates that this is not required in general.

Example 3.2. *Let $q(\rho) = e^{\frac{1}{2}(\ln^2(\alpha) - \ln^2(\alpha\rho))}$ for some $\alpha \leq e^{-\frac{\sqrt{5}+1}{2}}$. Then $q : [0, 1] \rightarrow [0, 1]$ is an increasing convex bijection, and the risk premium is proportional to $\frac{q(\rho)}{q'(\rho)\rho} = \frac{1}{-\ln(\alpha\rho)} \xrightarrow{\rho \rightarrow 0} 0$.*

For this weighting function, the risk premium is made arbitrarily small by taking $\underline{x} \rightarrow -\infty$ and $P(\underline{x}) \rightarrow 0$ for a fixed level of optimism. This case is more appealing than Example 3.1, because extreme optimism over the entire distribution is not required to approximate first best allocations. Indeed, extreme optimism actually causes the agent to exhibit risk seeking behavior in the RDEU framework, which entices the principal away from offering an approximately risk free contract in the first place. In contrast, extreme optimism in the lower tail is what makes a penalty contract so attractive because the agent is minimally risk averse, or even risk seeking, in this region.

Conclusion

While I have not allowed it here, an interesting feature of the RDEU framework is that an agent who exhibits declining marginal utility of wealth can actually be risk seeking if he is sufficiently optimistic. Moreover, in the RDEU framework extreme optimism and extreme overconfidence are one and the same as the agent increasingly overweights the probability of the best outcome. This feature potentially rationalizes several opposing empirical findings in the literature on optimism and overconfidence. For example, Humphery-Jenner et al. (2016) find evidence that overconfident executives, which they carefully distinguish from Otto's (2014) optimistic CEOs, receive *more* incentive-heavy compensation contracts. Given the opposing findings from these

two studies, the distinction between overconfidence and optimism may be more than semantic. When modeled through RDEU, it may be equivalent to the distinction between risk seeking and risk averse preferences given an optimistic weighting function.

Additionally, Ahmed and Duellman (2013) provide empirical evidence that overconfident CEO's are less likely to apply conservative accounting methods, both conditional and unconditional. Hilary et al. (2016) find somewhat inconsistent results with respect to optimism. They state that "measures of accruals or real earnings management are not affected by this over-optimism," and suggest that "managers, being over-optimistic regarding the likelihood of meeting the expectations they set, do not feel the need to manage earnings in order to reach their forecasts." Again, if overconfidence leads to risk seeking behavior as suggested by the RDEU framework, then the risk-free features of penalty contracts cease to be desirable and the optimality of conservatism wanes.

I have shown that unconditional conservatism enables the principal to game the manager's optimistic behavioral profile. By offering a contract that focuses in on the behavioral quirk, the principal can induce productive activity while settling up with a relatively small risk premium. The required degree of noise and bias to pull off the result is extreme. If the fundamental outcomes are evenly spaced, then

$$(\gamma_N, \gamma_{N-1}, \gamma_{N-2}, \dots, \gamma_1, \gamma_0) = \left(0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \dots, \frac{N-1}{N}, \frac{\pi_N - \pi_0}{\pi_N - \underline{x}}\right).$$

Only γ_0 approaches zero as $\underline{x} \rightarrow -\infty$, whereas the majority of the conservative parameters are fixed very close to one provided N is large. Such extreme bias is not dissimilar to immediate expensing: an accounting report of x_0 , which is very likely given this accounting system, implies that the asset is not capitalized at all.

References

- Ahmed, Anwer S, and Scott Duellman. 2013. Managerial overconfidence and accounting conservatism. *Journal of Accounting Research* 51 (1): 1–30.
- Arya, Anil, Jonathan Glover, and Brian Mittendorf. 2007. Aligning incentives by capping bonuses. In *Essays in Accounting Theory in Honour of Joel S. Demski*, 169–182. Springer.
- Ball, Ray, and Lakshmanan Shivakumar. 2005. Earnings quality in UK private firms: Comparative loss recognition timeliness. *Journal of Accounting and Economics* 39 (1): 83–128.
- Beaver, William H, Maureen F McNichols, and Karen K Nelson. 2007. An alternative interpretation of the discontinuity in earnings distributions. *Review of Accounting Studies* 12 (4): 525–556.
- Bertomeu, Jeremy. 2008. Risk management and the design of efficient incentive schemes. Dissertation, Carnegie Mellon University.
- Bushman, Robert M, Joseph D Piotroski, and Abbie J Smith. 2011. Capital allocation and timely accounting recognition of economic losses. *Journal of Business Finance & Accounting* 38 (1-2): 1–33.
- Dichev, Ilia D, John R Graham, Campbell R Harvey, and Shiva Rajgopal. 2013. Earnings quality: evidence from the field. *Journal of Accounting and Economics* 56 (2): 1–33.
- Durtschi, Cindy, and Peter Easton. 2005. Earnings management? the shapes of the frequency distributions of earnings metrics are not evidence ipso facto. *Journal of Accounting Research* 43 (4): 557–592.
- FASB. 2010. Conceptual framework for financial reporting. *Statement of Financial Accounting Concepts No. 8*.
- Friedman, Milton. 1953. The methodology of positive economics.
- Gigler, Frank B, and Thomas Hemmer. 2001. Conservatism, optimal disclosure policy, and the timeliness of financial reports. *The Accounting Review* 76 (4): 471–493.
- Gilboa, Itzhak. 1987. Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics* 16 (1): 65–88.
- Gu, Zhaoyang, and Joanna Shuang Wu. 2003. Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics* 35 (1): 5–29.
- Hellwig, Martin F. 2007. The role of boundary solutions in principal–agent problems of the Holmström–Milgrom type. *Journal of Economic Theory* 136 (1): 446–475.
- Hemmer, Thomas, and Eva Labro. 2008. On the optimal relation between the properties of managerial and financial reporting systems. *Journal of Accounting Research* 46 (5): 1209–1240.

- Hilary, Gilles, Charles Hsu, Benjamin Segal, and Rencheng Wang. 2016. The bright side of managerial over-optimism. *Journal of Accounting and Economics* 62 (1): 46–64.
- Hirshleifer, David, Angie Low, and Siew Hong Teoh. 2012. Are overconfident CEOs better innovators? *The Journal of Finance* 67 (4): 1457–1498.
- Holmström, Bengt. 1979. Moral hazard and observability. *The Bell Journal of Economics* 10 (1): 74–91.
- Holmstrom, Bengt, and Paul Milgrom. 1987. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society* 55 (2): 303–328.
- Holthausen, Robert W, David F Larcker, and Richard G Sloan. 1995. Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics* 19 (1): 29–74.
- Humphery-Jenner, Mark, Ling Lei Lisic, Vikram Nanda, and Sabatino Dino Silveri. 2016. Executive overconfidence and compensation structure. *Journal of Financial Economics* 119 (3): 533–558.
- Machina, Mark J. 1994. Review of ‘generalized expected utility theory: The rank-dependent model’. *Journal of Economic Literature* 32 (3): 1237–1238.
- Malmendier, Ulrike, and Geoffrey Tate. 2015. Behavioral CEOs: The role of managerial overconfidence. *The Journal of Economic Perspectives* 29 (4): 37–60.
- Mirrlees, James. 1974. Notes on welfare economics, information and uncertainty. *Essays on economic behavior under uncertainty*: 243–261.
- Murphy, Kevin J. 1999. Executive compensation. *Handbook of Labor Economics* 3:2485–2563.
- . 2013. Executive compensation: Where we are, and how we got there. *Handbook of the Economics of Finance: Corporate Finance* 2:211–356.
- Murphy, Kevin J, and Michael C Jensen. 2011. CEO bonus plans: And how to fix them. *Harvard Business School NOM Unit Working Paper*: 12–022.
- Otto, Clemens A. 2014. CEO optimism and incentive compensation. *Journal of Financial Economics* 114 (2): 366–404.
- PCAOB. 2010. Auditing standards related to the auditor’s assessment of and response to risk. *PCAOB Release No. 2010-004*.
- Quiggin, John. 1982. A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3 (4): 323–343.
- . 1992. *Generalized expected utility theory: The rank dependent model*. Springer Science & Business Media.
- Rogerson, William P. 1985. The first-order approach to principal-agent problems. *Econometrica: Journal of the Econometric Society*: 1357–1367.
- Schmeidler, David. 1989. Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*: 571–587.

- Segal, Uzi. 1987. The ellsberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*: 175–202.
- . 1990. Two-stage lotteries without the reduction axiom. *Econometrica: Journal of the Econometric Society*: 349–377.
- Shin, Hyun Song. 1994. News management and the value of firms. *The RAND Journal of Economics* 25 (1): 58–71.
- Starmer, Chris. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38 (2): 332–382.
- Sung, Jaeyoung. 1995. Linearity with project selection and controllable diffusion rate in continuous-time principal-agent problems. *The RAND Journal of Economics*: 720–743.
- Wakker, Peter. 1990. Under stochastic dominance choquet-expected utility and anticipated utility are identical. *Theory and Decision* 29 (2): 119–132.
- Watts, Ross L. 2003. Conservatism in accounting part I: Explanations and implications. *Accounting Horizons* 17 (3): 207–221.

Appendix A

Proofs

Proof of Proposition 1.1: For $k \in \{m_j, \dots, n_j\}$, the sufficient statistic assumption implies that $P(x_i)$ is linear in p , which yields $\frac{\partial P(x_i)}{\partial p_k} = \Pr(x_i|\pi_k)$ by the law of total probability. It follows that

$$\frac{\partial E[v]}{\partial p_k} = \sum_{i=0}^N \frac{\partial P(x_i)}{\partial p_k} v_i - c_k = \sum_{i=m_j}^{n_j} \Pr(x_i|\pi_k)(v_i - v_0) - c_k = 0. \quad (\text{A.1})$$

Now the N first order conditions can be expressed in matrix notation $\mathbf{b} = \mathbf{A}\mathbf{x}$:

$$\begin{bmatrix} \nabla^1 \\ \nabla^2 \\ \vdots \\ \nabla^{J-1} \\ \nabla^J \end{bmatrix} = \begin{bmatrix} \mathbf{P}^1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}^{J-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{P}^J \end{bmatrix} \begin{bmatrix} \mathbf{v}^1 \\ \mathbf{v}^2 \\ \vdots \\ \mathbf{v}^{J-1} \\ \mathbf{v}^J \end{bmatrix},$$

where

$$\nabla^j \equiv \begin{bmatrix} c_{m_j} \\ c_{m_j+1} \\ \vdots \\ c_{n_j-1} \\ c_{n_j} \end{bmatrix}, \quad \mathbf{P}^j \equiv \begin{bmatrix} \Pr(x_{m_j}|\pi_{m_j}) & \cdots & \Pr(x_{n_j}|\pi_{m_j}) \\ \vdots & \ddots & \vdots \\ \Pr(x_{m_j}|\pi_{n_j}) & \cdots & \Pr(x_{n_j}|\pi_{n_j}) \end{bmatrix}, \quad \text{and } \mathbf{v}^j \equiv \begin{bmatrix} v_{m_j} - v_0 \\ v_{m_j+1} - v_0 \\ \vdots \\ v_{n_j-1} - v_0 \\ v_{n_j} - v_0 \end{bmatrix}.$$

Since \mathbf{A} is a block diagonal matrix, I can treat this large system of equations as J independent systems written $\nabla^j = \mathbf{P}^j \mathbf{v}^j$. As long as \mathbf{P}^j is of full rank for all j , there exist unique utilities v_0, \dots, v_N that induce action p and bind the individual rationality constraint. \square

Proof of Proposition 1.2: Dropping the subscripts, the j^{th} system of equations can be written

$$\begin{bmatrix} c_m \\ c_{m+1} \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} 1-\theta_m & (1-\theta_{m+1})\delta_m & \cdots & (1-\theta_{n-1})\delta_m \prod_{i=m+1}^{n-2} \theta_i & (1-\theta_n)\delta_m \prod_{i=m+1}^{n-1} \theta_i \\ (1-\theta_m)\gamma_{m+1} & 1-\theta_{m+1} & \cdots & (1-\theta_{n-1})\delta_{m+1} \prod_{i=m+2}^{n-2} \theta_i & (1-\theta_n)\delta_{m+1} \prod_{i=m+2}^{n-1} \theta_i \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (1-\theta_m)\gamma_{n-1} \prod_{i=m+1}^{n-2} \theta_i & (1-\theta_{m+1})\gamma_{n-1} \prod_{i=m+2}^{n-2} \theta_i & \cdots & 1-\theta_{n-1} & (1-\theta_n)\delta_{n-1} \\ (1-\theta_m)\gamma_n \prod_{i=m+1}^{n-1} \theta_i & (1-\theta_{m+1})\gamma_n \prod_{i=m+2}^{n-1} \theta_i & \cdots & (1-\theta_{n-1})\gamma_n & 1-\theta_n \end{bmatrix} \begin{bmatrix} v_m - v_0 \\ v_{m+1} - v_0 \\ \vdots \\ v_{n-1} - v_0 \\ v_n - v_0 \end{bmatrix}.$$

Setting $\theta_i = 1$ for all $i \in \{m, \dots, n\}$ yields

$$\begin{bmatrix} c_m \\ c_{m+1} \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} \gamma_m & 0 & \cdots & 0 & \delta_m \\ \gamma_{m+1} & 0 & \cdots & 0 & \delta_{m+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n-1} & 0 & \cdots & 0 & \delta_{n-1} \\ \gamma_n & 0 & \cdots & 0 & \delta_n \end{bmatrix} \begin{bmatrix} v_m - v_0 \\ v_{m+1} - v_0 \\ \vdots \\ v_{n-1} - v_0 \\ v_n - v_0 \end{bmatrix}.$$

Equation (1.6) immediately follows.

Suppose that c_k is increasing convex in k so that $\frac{c_k - c_m}{c_n - c_m} < \frac{\pi_k - \pi_m}{\pi_n - \pi_m}$. Substituting $\delta_m = \gamma_n = 0$ into (1.6) yields $c_i = \gamma_i c_m + \delta_i c_n$. These observations imply that

$$\pi_n \delta_k + \pi_m \gamma_k = (\pi_n - \pi_m) \frac{c_k - c_m}{c_n - c_m} + \pi_m < (\pi_n - \pi_m) \frac{\pi_k - \pi_m}{\pi_n - \pi_m} + \pi_m = \pi_k,$$

which implies that

$$\sum_{i=m}^n \pi_i P(x_i) = \sum_{i=m}^n (\pi_n \delta_i + \pi_m \gamma_i) p_i < \sum_{i=m}^n \pi_i p_i.$$

That is, the optimal system is conservative over $\{\pi_m, \dots, \pi_n\}$. Symmetrically, if c_k is increasing concave (linear) in k then the optimal system is aggressive (unbiased) over $\{\pi_m, \dots, \pi_n\}$. \square

Proof of Proposition 1.3: First, it can be shown that $\sum_{i=0}^N v_i P(x_i) = v_0 + \sum_{i=1}^N c_i p_i$, which allows the individual rationality constraint to be written $v_0 + \sum_{i=1}^N c_i p_i - c(p) \geq \bar{v}$. With choice variables p , v_{m_j} , and v_{n_j} , and with multiplier λ on the individual rationality constraint, the Lagrangian is given by

$$\begin{aligned} \mathcal{L} = & \sum_{i=0}^N \pi_i p_i - \sum_{i=1}^N h(v_i) P(x_i) + \lambda \left(v_0 + \sum_{i=1}^N c_i p_i - c(p) - \bar{v} \right) \\ & + \sum_{j=1}^J \sum_{i=m_j}^{n_j} \mu_i (c_i - v_{m_j} + v_0) + \sum_{j=1}^J \sum_{i=m_j}^{n_j} \nu_i (v_{n_j} - v_0 - c_i). \end{aligned} \quad (\text{A.2})$$

Substituting (1.7) into δ_i and γ_i for each $i \in \{m, \dots, n\}$ yields

$$\begin{aligned} \frac{\partial P(x_m)}{\partial p_k} &= \frac{\partial}{\partial p_k} \left(\sum_{i=m}^n \frac{v_n - v_0 - c_i}{v_n - v_m} p_i \right) = \gamma_k - \frac{c_{kk} p_k}{v_n - v_m} \\ \frac{\partial P(x_n)}{\partial p_k} &= \frac{\partial}{\partial p_k} \left(\sum_{i=m}^n \frac{c_i - v_m + v_0}{v_n - v_m} p_i \right) = \delta_k + \frac{c_{kk} p_k}{v_n - v_m} \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} \frac{\partial P(x_m)}{\partial v_m} &= \frac{\partial}{\partial v_m} \left(\sum_{i=m}^n \frac{v_n - v_0 - c_i}{v_n - v_m} p_i \right) = \sum_{i=m}^n \frac{v_n - v_0 - c_i}{(v_n - v_m)^2} p_i = \frac{P(x_m)}{v_n - v_m} \\ \frac{\partial P(x_m)}{\partial v_n} &= \frac{\partial}{\partial v_n} \left(\sum_{i=m}^n \frac{v_n - v_0 - c_i}{v_n - v_m} p_i \right) = \sum_{i=m}^n \frac{c_i - v_m + v_0}{(v_n - v_m)^2} p_i = \frac{P(x_n)}{v_n - v_m} \\ \frac{\partial P(x_n)}{\partial v_m} &= \frac{\partial}{\partial v_m} \left(\sum_{i=m}^n \frac{c_i - v_m + v_0}{v_n - v_m} p_i \right) = \sum_{i=m}^n \frac{c_i - v_n + v_0}{(v_n - v_m)^2} p_i = \frac{-P(x_m)}{v_n - v_m} \\ \frac{\partial P(x_n)}{\partial v_n} &= \frac{\partial}{\partial v_n} \left(\sum_{i=m}^n \frac{c_i - v_m + v_0}{v_n - v_m} p_i \right) = \sum_{i=m}^n \frac{v_m - v_0 - c_i}{(v_n - v_m)^2} p_i = \frac{-P(x_n)}{v_n - v_m}, \end{aligned} \quad (\text{A.4})$$

where $m \in \{m_1, \dots, m_J\}$ and $n \in \{n_1, \dots, n_J\}$. Using (A.3) and (A.4) in the first order conditions of (A.2) with respect to p_i , v_m , and v_n yields the result.

Finally, (1.10) strictly positive implies that $\delta_k = \gamma_l = 1$ for at least one k and l between m and n . It follows immediately that $\delta_n = \gamma_m = 1$ as long as c_i is nondecreasing in i . \square

Proof of Lemma 1.1: Suppose for the moment that only μ_m and ν_n are positive. Then $c_n = v_n - v_0$ and $c_m = v_m - v_0$. Substituting these into δ_i and γ_i yields the following Lagrangian:

$$\begin{aligned} \mathcal{L} = & \sum_{i=0}^N \pi_i p_i - \sum_{j=1}^J h(v_0 + c_{n_j}) P(x_{n_j}) - \sum_{j=1}^J h(v_0 + c_{m_j}) P(x_{m_j}) - h(v_0) p_0 \\ & + \lambda \left(v_0 + \sum_{i=1}^N c_i p_i - c(p) - \bar{v} \right). \end{aligned} \quad (\text{A.5})$$

Suppressing the j subscripts, it can be shown that

$$\frac{\partial P(x_m)}{\partial p_k} = \frac{\partial}{\partial p_k} \left(\sum_{i=m}^n \frac{c_n - c_i}{c_n - c_m} p_i \right) = \begin{cases} 1 + \frac{c_{mm}}{c_n - c_m} \sum_{i=m+1}^{n-1} \gamma_i p_i & \text{if } k = m \\ \frac{c_{nn}}{c_n - c_m} \sum_{i=m+1}^{n-1} \delta_i p_i & \text{if } k = n, \end{cases} \quad (\text{A.6})$$

$$\frac{\partial P(x_n)}{\partial p_k} = \frac{\partial}{\partial p_k} \left(\sum_{i=m}^n \frac{c_i - c_m}{c_n - c_m} p_i \right) = \begin{cases} -\frac{c_{mm}}{c_n - c_m} \sum_{i=m+1}^{n-1} \gamma_i p_i & \text{if } k = m \\ 1 - \frac{c_{nn}}{c_n - c_m} \sum_{i=m+1}^{n-1} \delta_i p_i & \text{if } k = n. \end{cases}$$

The first order conditions of (A.5) with respect to p_n and p_m are respectively given by

$$\begin{aligned} \pi_n - \pi_0 = & h(v_0 + c_n) - h(v_0) + (h'(v_0 + c_n) - \lambda) c_{nn} p_n \\ & + \left(h'(v_0 + c_n) - \frac{h(v_0 + c_n) - h(v_0 + c_m)}{c_n - c_m} \right) c_{nn} \sum_{i=m+1}^{n-1} \delta_i p_i \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} \pi_m - \pi_0 = & h(v_0 + c_m) - h(v_0) + (h'(v_0 + c_m) - \lambda) c_{mm} p_m \\ & - \left(\frac{h(v_0 + c_n) - h(v_0 + c_m)}{c_n - c_m} - h'(v_0 + c_m) \right) c_{mm} \sum_{i=m+1}^{n-1} \gamma_i p_i \end{aligned} \quad (\text{A.8})$$

Notice that the first lines of (A.7) and (A.8) are equivalent to (SB_{*i*}). The second lines of (A.7) and (A.8) are respectively positive and negative, implying that c_n (c_m) must be chosen smaller (larger) in the third best than in the second best to ensure that the equalities hold. \square

Proof of Lemma 1.2: Let \underline{h}_i and \bar{h}_i represent the expected dollar wages corresponding to \underline{v}_i and \bar{v}_i . Fixing p and $\theta_i \in (0, 1)$ for all i , the principal chooses δ_i and v_i to minimize the following program:

$$\begin{aligned} \min_{v_i, \delta_i} \quad & \underline{h}_i \gamma_i + h(v_i)(1 - \theta_i) + \bar{h}_i \delta_i \\ \text{s.t.} \quad & c_i = \underline{v}_i \gamma_i + v_i(1 - \theta_i) + \bar{v}_i \delta_i - v_0. \end{aligned} \tag{A.9}$$

That is, the principal chooses v_i and δ_i , with $\gamma_i \equiv \theta_i - \delta_i$, to minimize the expected dollar wage while satisfying the incentive compatibility constraint obtained from (A.1). Notice that if δ_i and v_i are varied while continuing to meet the constraint in (A.9), the agent's chosen action and expected utility remain unchanged, so the individual rationality constraint is automatically satisfied.¹ Solving the constraint for δ_i and substituting into the objective function yields

$$\min_{v_i} \quad h(v_i)(1 - \theta_i) + (\bar{h}_i - \underline{h}_i) \frac{c_i - \underline{v}_i \theta_i + v_0 - v_i(1 - \theta_i)}{\bar{v}_i - \underline{v}_i} + \underline{h}_i \theta_i.$$

The first order condition is given by

$$h'(v_i)(1 - \theta_i) - \frac{\bar{h}_i - \underline{h}_i}{\bar{v}_i - \underline{v}_i}(1 - \theta_i) = 0.$$

The second order condition, $h''(v_i)(1 - \theta_i) > 0$, reveals that this is indeed characterizes a minimum. Thus the principal selects v_i (and δ_i) so that $h'(v_i) = \frac{\bar{h}_i - \underline{h}_i}{\bar{v}_i - \underline{v}_i}$. By the convexity of $h(\cdot)$, v_i lies strictly between \underline{v}_i and \bar{v}_i . \square

¹The agent's cost of effort, $c(p)$, is held constant by assumption, whereas his expected wage in utiles, $E[v] = \sum_i E[v|\pi_i]p_i$, is constant by the right hand side of the constraint in (A.9).

Proof of Lemma 1.3: The constraint in (A.9) can be rewritten

$$v_i = v_0 + c_i + \frac{c_i - (v_i - v_0)}{1 - \theta_i} \gamma_i + \frac{c_i - (\bar{v}_i - v_0)}{1 - \theta_i} \delta_i. \quad (\text{A.10})$$

I begin with the capped region, which satisfies $c_i = v_n - v_0$. Note that $\bar{v}_{n-1} = v_n$ by definition.

Thus for $i = n - 1$, (A.10) reduces to

$$v_{n-1} = v_n + \frac{v_n - v_{n-1}}{1 - \theta_{n-1}} \gamma_{n-1} + \frac{v_n - \bar{v}_{n-1}}{1 - \theta_{n-1}} \delta_{n-1} = v_n + \frac{v_n - v_{n-1}}{1 - \theta_{n-1}} \gamma_{n-1} \geq v_n,$$

with a strict inequality if $\gamma_{n-1} > 0$. But $v_{n-1} > v_n = \bar{v}_{n-1}$ contradicts Lemma 1.2, implying that γ_{n-1} must equal zero. Thus $v_{n-1} = v_n$, which implies that $\bar{v}_{n-2} = v_n$. Let the inductive hypothesis be that $\bar{v}_i = v_n$ in the capped region. Then (A.10) reduces to

$$v_i = v_n + \frac{v_n - v_i}{1 - \theta_i} \gamma_i + \frac{v_n - \bar{v}_i}{1 - \theta_i} \delta_i = v_n + \frac{v_n - v_i}{1 - \theta_i} \gamma_i \geq v_n,$$

with a strict inequality if $\gamma_i > 0$. But $v_{n-1} > v_n = \bar{v}_{n-1}$ contradicts Lemma 1.2, implying that γ_i must equal zero. Thus $v_i = v_{i+1} = \dots = v_n$, which implies that $\bar{v}_{i-1} = v_n$. By the principle of mathematical induction, if c_i is capped then so is v_i .

The proof that v_i exhibits a floor whenever c_i exhibits a floor is perfectly symmetric to the proof for a cap. □

Proof of Proposition 1.4: Fixing $\nu_i = \mu_i = 0$ for all i and recalling from Proposition 1.3 that $c_{m_j} = v_{m_j} - v_0$, (TB _{i}) can be rewritten

$$\pi_i - \pi_0 = \frac{h(v_{n_j}) - h(v_{m_j})}{v_{n_j} - v_{m_j}} (c_i - c_{m_j}) + h(v_{m_j}) - h(v_0) + \left(\frac{h(v_{n_j}) - h(v_{m_j})}{v_{n_j} - v_{m_j}} - \lambda \right) c_{ii} p_i.$$

Define $\alpha^j \equiv \frac{h(v_{n_j}) - h(v_{m_j})}{v_{n_j} - v_{m_j}}$. Subtracting (TB _{m_j}) from (TB _{i}) yields

$$\pi_i - \pi_{m_j} = \alpha^j (c_i - c_{m_j}) + (\alpha^j - \lambda) (c_{ii} p_i - c_{m_j m_j} p_{m_j}). \quad (\text{A.11})$$

Choose j^* such that $\alpha^j - \lambda$ is positive (negative) for all $j \geq j^*$ ($j < j^*$).²

Case 1: $j \geq j^*$. The left hand side of (A.11) is linearly increasing in i . Since $\alpha^j - \lambda > 0$, if $c_{ii} p_i$ is concave (convex) in i for all $j \geq j^*$, then c_i must be convex (concave) in i .

Case 2: $j < j^*$. The left hand side of (A.11) is linearly increasing in i . Since $\alpha^j - \lambda < 0$, if $c_{ii} p_i$ is concave (convex) in i for all $j < j^*$, then c_i must be concave (convex) in i .

Case 3: $c_{ii} p_i = a^j c_i$. Then (A.11) can be rewritten $\pi_i - \pi_{m_j} = (\alpha^j + (\alpha^j - \lambda) a^j) (c_i - c_{m_j})$, which implies that c_i is linear in i over $\{\pi_{m_j}, \dots, \pi_{n_j}\}$. \square

²This is always possible since α^j is the slope of nonoverlapping secant lines connecting increasing points on a convex function and is therefore increasing in j , whereas λ is a weighted average of $h'(v_0)$ and $\{\alpha^j\}_{j=1}^J$. To see this, define $q_j \equiv \sum_{i=m_j}^{n_j} p_i$. Then the first order condition with respect to v_0 reveals that

$$\lambda = h'(v_0) p_0 + \sum_{j=1}^J \alpha^j q_j + \sum_{j=1}^J \sum_{i=m_j}^{n_j} (\nu_i - \mu_i). \quad (\text{A.12})$$

Proof of Proposition 2.2: Recall that the agent is risk neutral, and let the principal take γ_i and δ_i as exogenous. Then the principal chooses v_{m_j} and v_{n_j} to maximize the following program:

$$\begin{aligned} \max_{v_{m_j}, v_{n_j}} \quad & \pi_0 + \sum_{i=1}^N (\phi_i^k + \pi_i - \pi_0) p_i - \bar{v} - c(p) \\ \text{s.t.} \quad & c_i = \gamma_i v_{m_j} + \delta_i v_{n_j} - v_0 \quad \text{for all } i \in \{m_j, \dots, n_j\} \text{ and } j \in \{1, \dots, J\}, \end{aligned}$$

where the individual rationality constraint is substituted directly into the objective function since the agent is risk neutral. Implicitly substitute the incentive compatibility constraints into the objective function so that p_i is an implicit function of v_{m_j} and v_{n_j} . Suppressing the j subscripts, apply the chain rule and differentiate the incentive compatibility constraint to obtain $\frac{\partial p_i}{\partial v_m} = \frac{\partial p_i}{\partial c_i} \frac{\partial c_i}{\partial v_m} = \frac{\gamma_i}{c_{ii}}$ and $\frac{\partial p_i}{\partial v_n} = \frac{\partial p_i}{\partial c_i} \frac{\partial c_i}{\partial v_n} = \frac{\delta_i}{c_{ii}}$. Using this and the total derivative of $c(p)$, the first order conditions for v_m and v_n are respectively given by

$$0 = \sum_{i=m}^{n-1} (\phi_i^k + \pi_i - \pi_0 - c_i) \frac{\gamma_i}{c_{ii}} \quad \text{and} \quad 0 = \sum_{i=m+1}^n (\phi_i^k + \pi_i - \pi_0 - c_i) \frac{\delta_i}{c_{ii}}.$$

With little if any loss of generality, let $|\Pi^j| = 3$ and denote $l \equiv n - 1 = m + 1$. Noting that $\delta_m = \gamma_n = 0$, these first order conditions can be rearranged and combined to yield

$$\frac{\phi_m^k + \pi_m - \pi_0 - c_m}{\gamma_l c_{mm}} = \frac{c_l - (\phi_l^k + \pi_l - \pi_0)}{c_{ll}} = \frac{\phi_n^k + \pi_n - \pi_0 - c_n}{\delta_l c_{nn}}. \quad (\text{A.13})$$

Suppose that γ_l and δ_l are chosen such that c_i is linear in i . Then for ϕ_i^k concave (convex) in i , (A.13) is negative (positive), which implies that c_m is greater than (less than) $\phi_m^k + \pi_m - \pi_0$. Recalling that (2.2) specifies $c_m = \phi_m^k + \pi_m - \pi_0$, the result immediately follows. \square

Proof of Proposition 3.1: It easily follows from (3.3) that $\underline{g}(\mathbf{P}) + \sum_{i=0}^N g_i(\mathbf{P}) = 1$. Recall that $\underline{g}(\mathbf{P}) = q(\mathbf{P}(\underline{x}))$. Rewriting the principal's program under the assumption that $v_i = v > \underline{v}$ for all i yields

$$\begin{aligned} \max_{p, v, \underline{v}} \quad & \sum_{i=0}^N \pi_i p_i - \sum_{i=0}^N h(v_i) \mathbf{P}(x_i) - h(\underline{v}) \mathbf{P}(\underline{x}) \\ \text{s.t.} \quad & v - q(\mathbf{P}(\underline{x}))(v - \underline{v}) - c(p) \geq \bar{v} \\ & p \in \operatorname{argmax}_{\tilde{p}} v - q(\mathbf{P}(\underline{x}))(v - \underline{v}) - c(\tilde{p}). \end{aligned}$$

It follows from (1.1) that $\mathbf{P}(\underline{x})$ is a linear function of p . Since composition with a linear mapping preserves convexity or concavity, $q \circ P$ is convex (concave) if q is convex (concave). If q and c are both convex, then the incentive compatibility constraint is a concave function of p , implying that the first order conditions are necessary and sufficient for an interior solution.

This result can be obtained explicitly. Using (1.2), the agent's first order condition with respect to p_k is given by

$$\frac{\partial EU}{\partial p_k} = -q'(\mathbf{P}(\underline{x}))(\Pr(\underline{x}|\pi_k) - \gamma_0)(v - \underline{v}) - c_k = 0, \quad (\text{A.14})$$

which yields (3.5), and the second partial derivatives can be written

$$\frac{\partial^2 EU}{\partial p_k \partial p_j} = -q''(\mathbf{P}(\underline{x}))(v - \underline{v})(\gamma_0 - \Pr(\underline{x}|\pi_k))(\gamma_0 - \Pr(\underline{x}|\pi_j)) - c_{kj}. \quad (\text{A.15})$$

Claim: The square matrix \mathbf{B} with kj^{th} entry $b_k b_j \equiv (\gamma_0 - \Pr(\underline{x}|\pi_k))(\gamma_0 - \Pr(\underline{x}|\pi_j))$ is positive semidefinite.

Proof of Claim: \mathbf{B} is positive semidefinite if and only if its eigenvalues are nonnegative. Notice that $\mathbf{B} = \mathbf{b}\mathbf{b}^T$, where

$$\mathbf{b}^T \equiv \begin{bmatrix} b_1 & b_2 & \cdots & b_{N-1} & b_N \end{bmatrix}.$$

For some nonzero eigenvector \mathbf{x} , the corresponding eigenvalue λ satisfies

$$\lambda \mathbf{x} = \mathbf{B}\mathbf{x} = \mathbf{b}\mathbf{b}^T \mathbf{x} \quad \implies \quad \lambda \mathbf{x}^T \mathbf{x} = \mathbf{x}^T \mathbf{b}\mathbf{b}^T \mathbf{x} = (\mathbf{b}^T \mathbf{x})^T \mathbf{b}^T \mathbf{x} \quad \implies \quad \lambda = \frac{\|\mathbf{b}^T \mathbf{x}\|^2}{\|\mathbf{x}\|^2} \geq 0.$$

■

If $q''(P(\underline{x}))$ is positive, the Claim implies that the Hessian matrix consisting of the first terms of each cross partial given by (A.15) is negative semidefinite. Since the sum of a negative semidefinite and a negative definite matrix is negative definite, the second order condition for a maximum is satisfied. The same cannot be said if $q''(P(\underline{x}))$ is negative, but since the second order conditions hold when $q''(P(\underline{x})) = 0$, continuity implies that they hold in some ϵ -neighborhood of zero.

□

Proof of Proposition 3.2: Solving the N^{th} first order condition given by (3.5) for $v - \underline{v}$ with $\gamma_N = 0$ yields $v - \underline{v} = \frac{c_N}{q'(P(\underline{x}))\gamma_0}$. Substituting this into the 1st first order condition given by (3.5) yields

$$1 - \gamma_1 = \frac{c_1}{c_N} \iff \gamma_1 = \frac{c_N - c_1}{c_N}.$$

Substituting this into the 2nd first order condition yields

$$1 - \gamma_2 \frac{c_N - c_1}{c_N} = \frac{c_2}{c_N} \iff \gamma_2 = \frac{c_N - c_2}{c_N} \frac{c_N}{c_N - c_1} = \frac{c_N - c_2}{c_N - c_1}.$$

Let the inductive hypothesis be that $\gamma_i = \frac{c_N - c_i}{c_N - c_{i-1}}$ for all $i < k$. Then the k^{th} first order condition can be written

$$1 - \gamma_k \prod_{i=1}^{k-1} \frac{c_N - c_i}{c_N - c_{i-1}} = 1 - \gamma_k \frac{c_N - c_{k-1}}{c_N} = \frac{c_k}{c_N} \iff \gamma_k = \frac{c_N - c_k}{c_N} \frac{c_N}{c_N - c_{k-1}} = \frac{c_N - c_k}{c_N - c_{k-1}}.$$

□

Proof of Proposition 3.3: Substituting (3.8) into the individual rationality constraint yields

$$v = \bar{v} + c(p) + q(P(\underline{x}))(v - \underline{v}) = \bar{v} + c(p) + \frac{q(P(\underline{x}))}{q'(P(\underline{x}))P(\underline{x})} \sum_{k=0}^N p_k(\pi_N - \pi_k). \quad (\text{A.16})$$

Since $\bar{v} + c(p)$ is paid when the agent's action is observable, the risk premium is equal to the last term. Moreover, $\sum_{k=0}^N p_k(\pi_N - \pi_k)$ is constant in \underline{x} given the first best action. It follows that the risk premium is proportional to $\frac{q(P(\underline{x}))}{q'(P(\underline{x}))P(\underline{x})}$.

□

Appendix B

Figures

Figure 1.1. Representation of the earnings process pioneered by Gigler and Hemmer (2001).

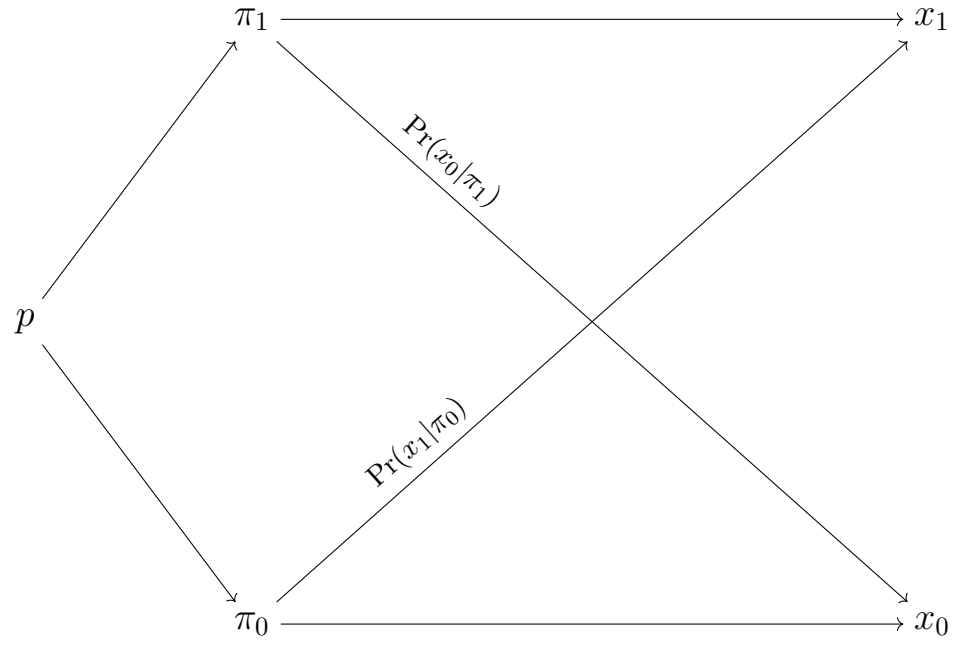


Figure 1.2. Unstructured extension of the Gigler and Hemmer earnings process with $N = 3$.

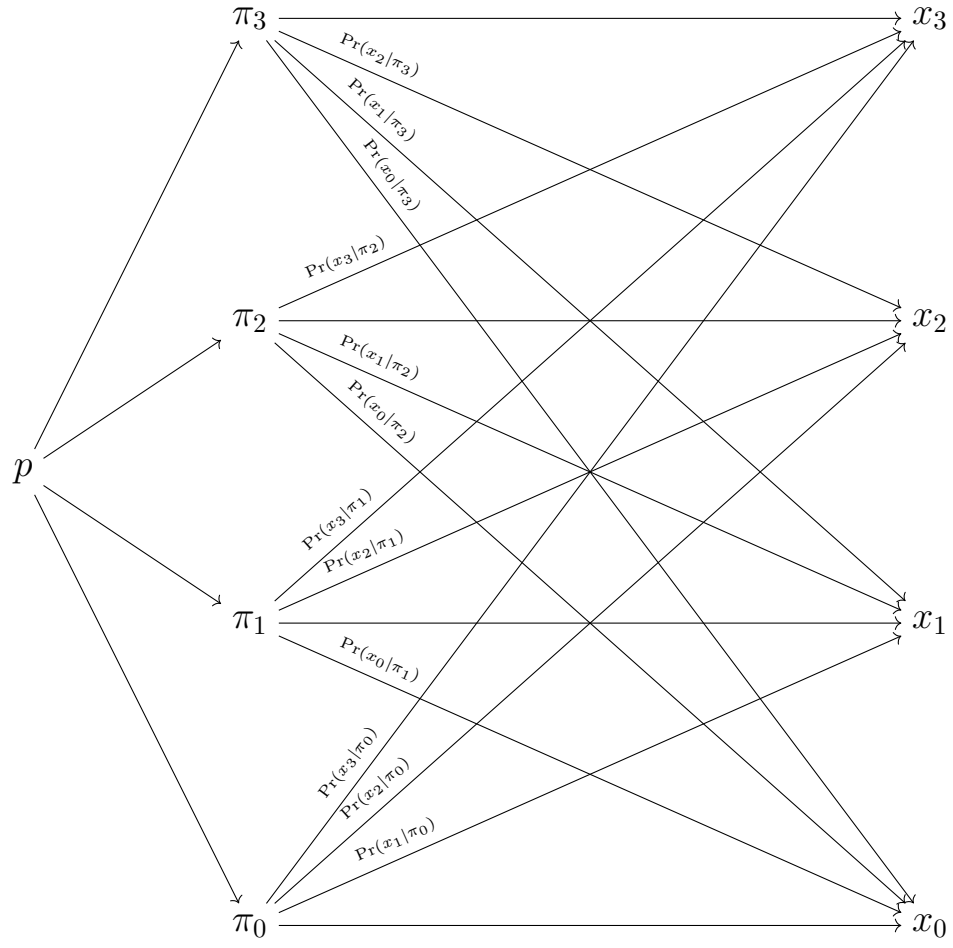


Figure 1.3. Structured extension of the Gigler and Hemmer earnings process with arbitrary N assuming that $\pi = \pi_i$ is realized.

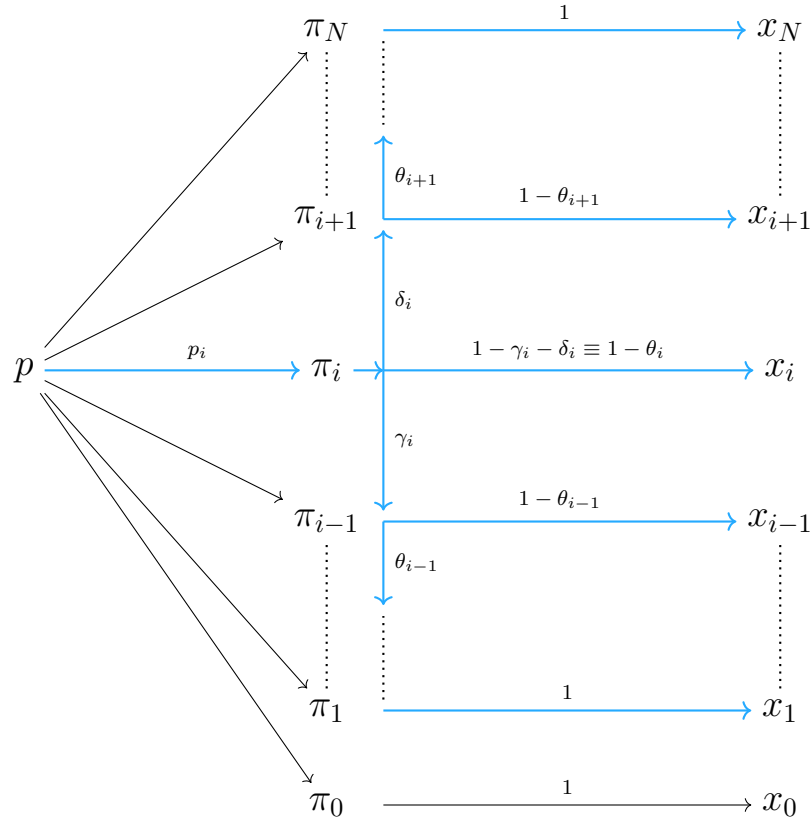


Figure 1.4. Illustration of the earnings process with J informationally distinct ranges, where $\Pi^j \equiv \{\pi_{m_j}, \dots, \pi_{n_j}\}$, $X^j \equiv \{x_{m_j}, \dots, x_{n_j}\}$, and $\Pr(x \in X^j | \pi \in \Pi^l) = 0$ if $j \neq l$.

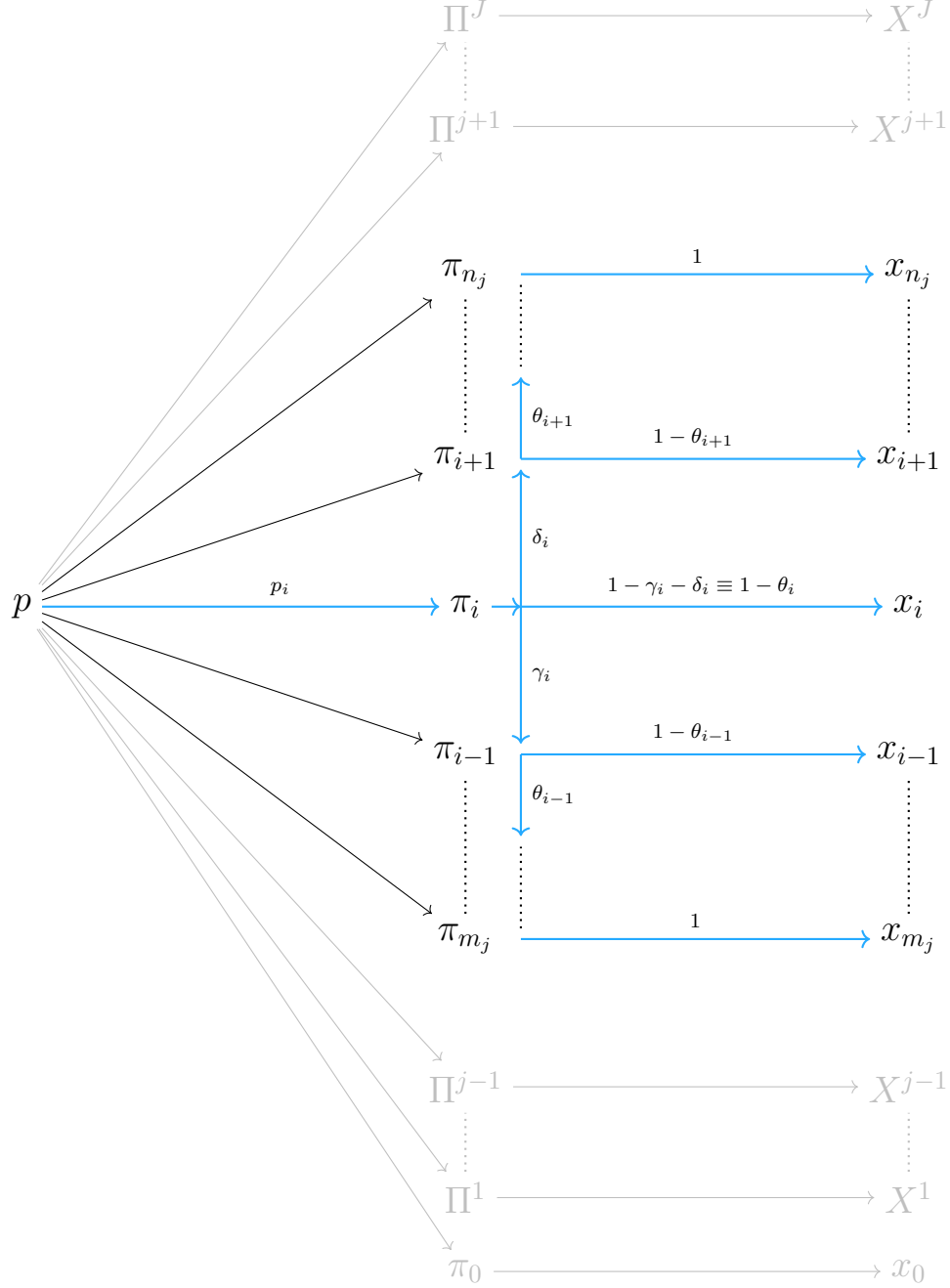


Figure 1.5. Illustration of the earnings process assuming that $\theta_i = 1$ for all $i \in \{m_j, \dots, n_j\}$, where $\gamma_{m_j} \equiv \Pr(x_{m_j} | \pi_{m_j})$ and $\delta_{n_j} \equiv \Pr(x_{n_j} | \pi_{n_j})$ for all $j \in \{1, \dots, J\}$.

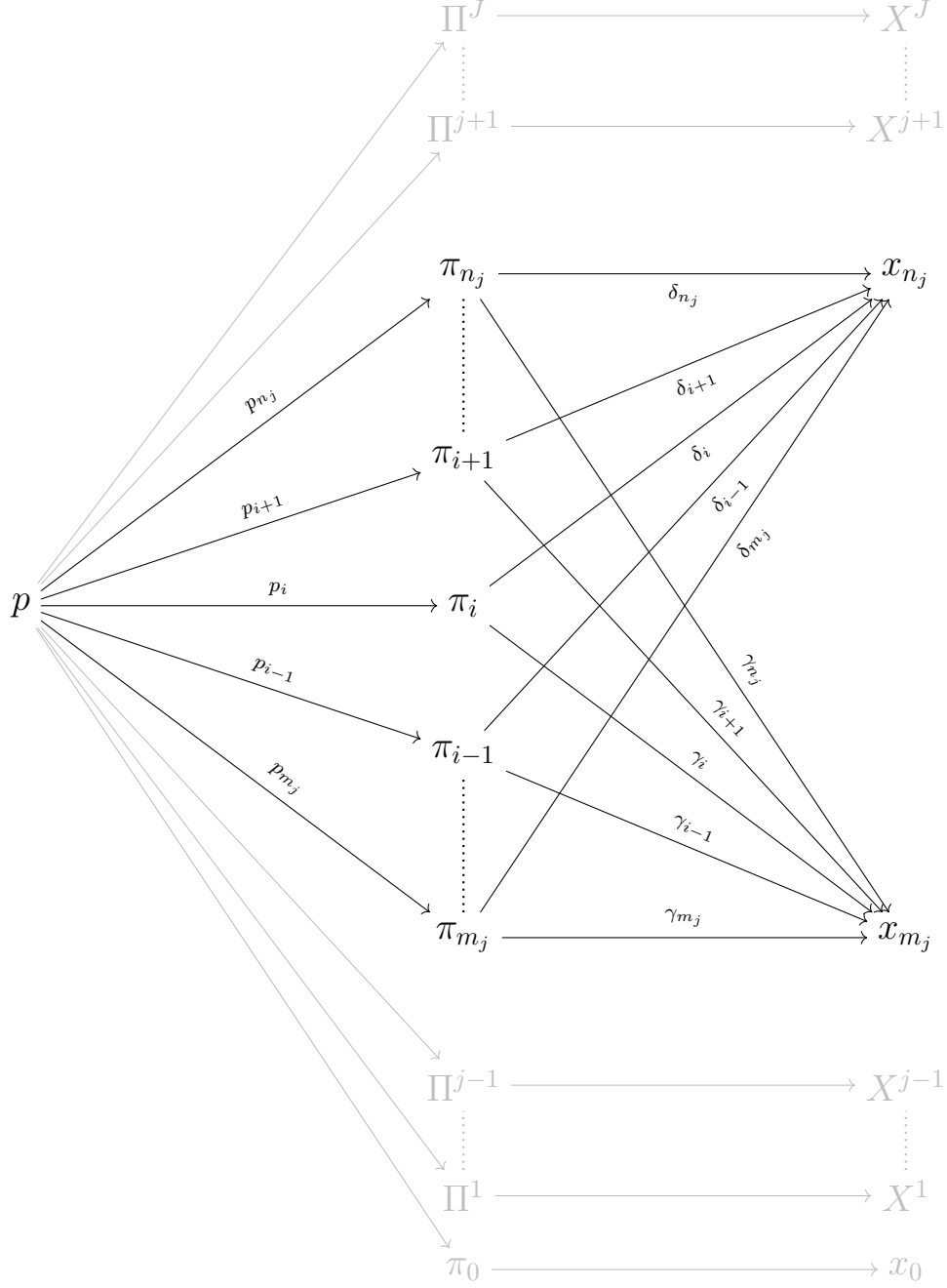


Figure 1.6. Correspondence between the convexity of c_i and the direction of bias.

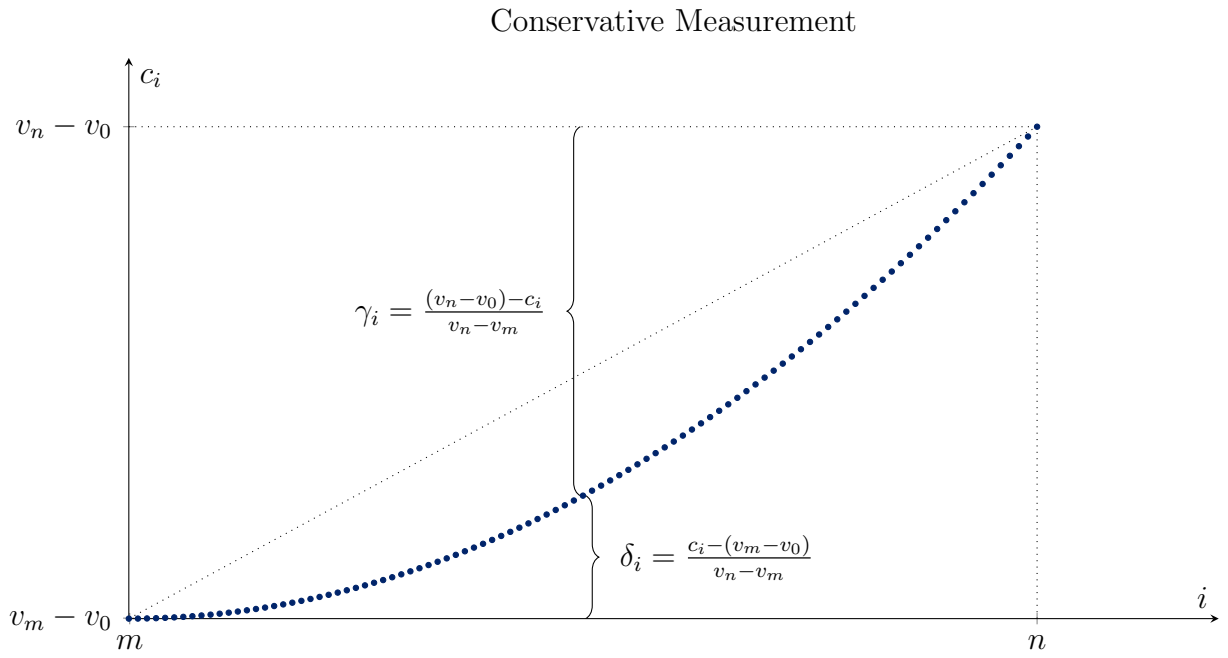
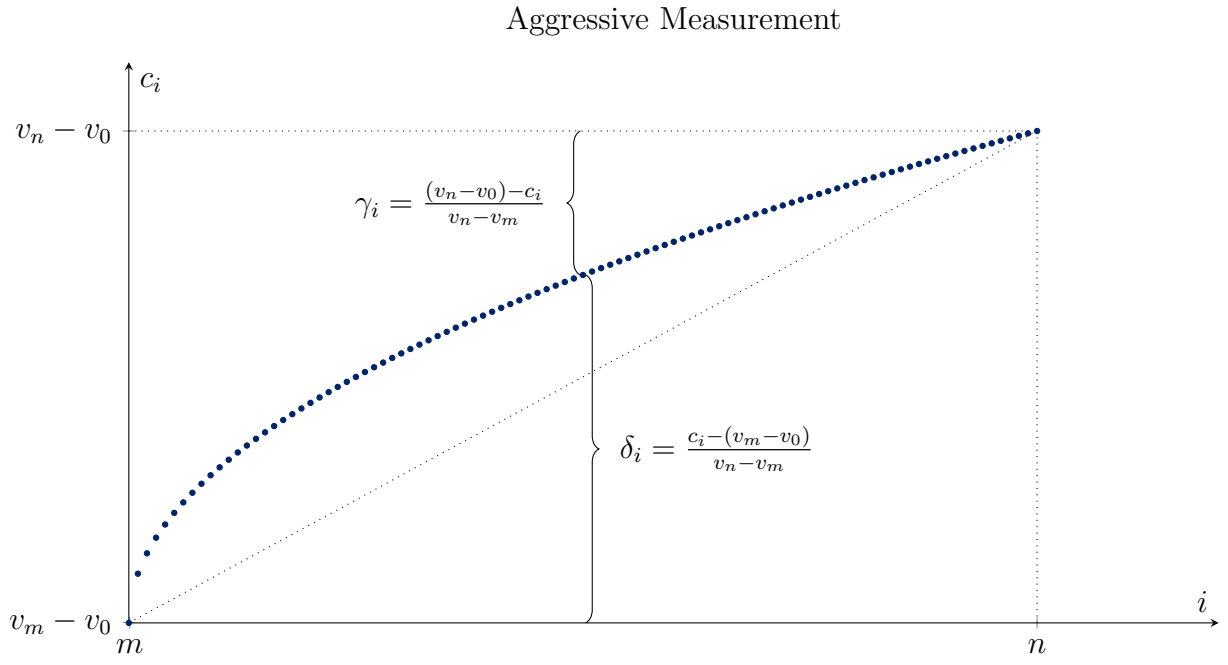


Figure 1.7. First best and hypothetical second best (no measurement error) marginal costs and fundamental earnings distributions.

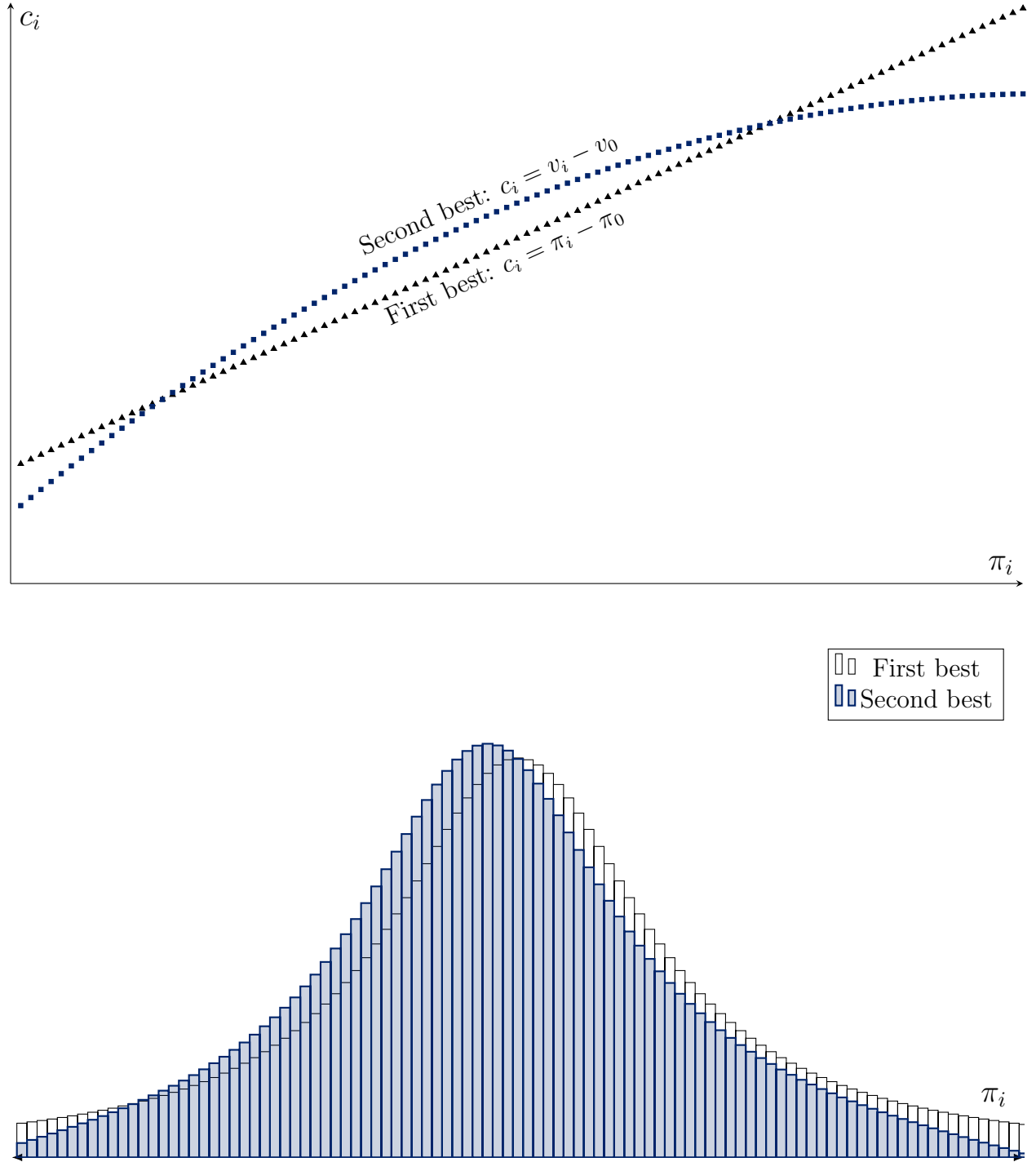


Figure 1.8. Second best (no measurement error) and third best (measurement error) marginal costs and fundamental earnings distributions with one informationally distinct range ($J = 1$).

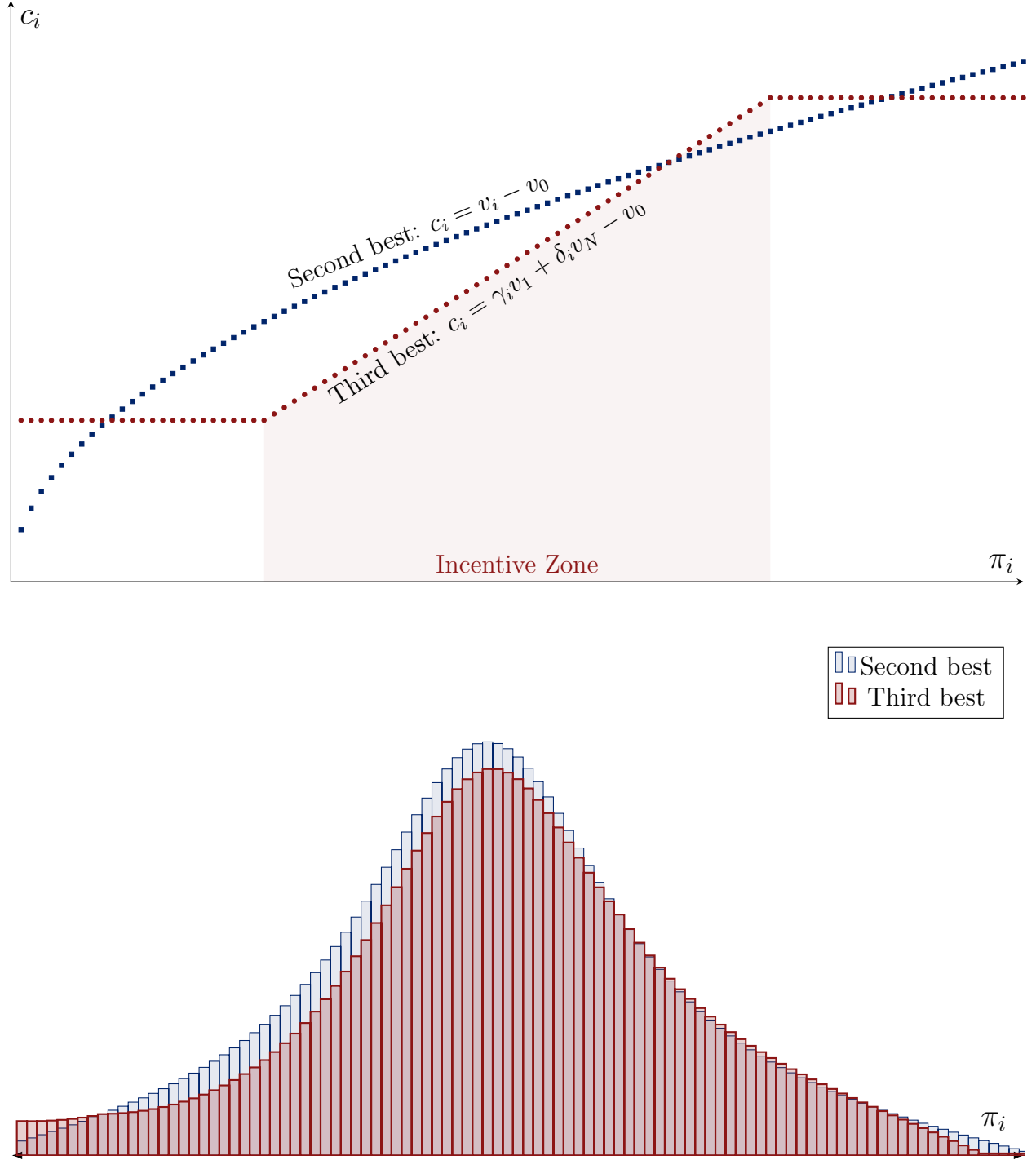


Figure 1.9. Implementing the third-best action through “all-or-nothing” measurement rules.

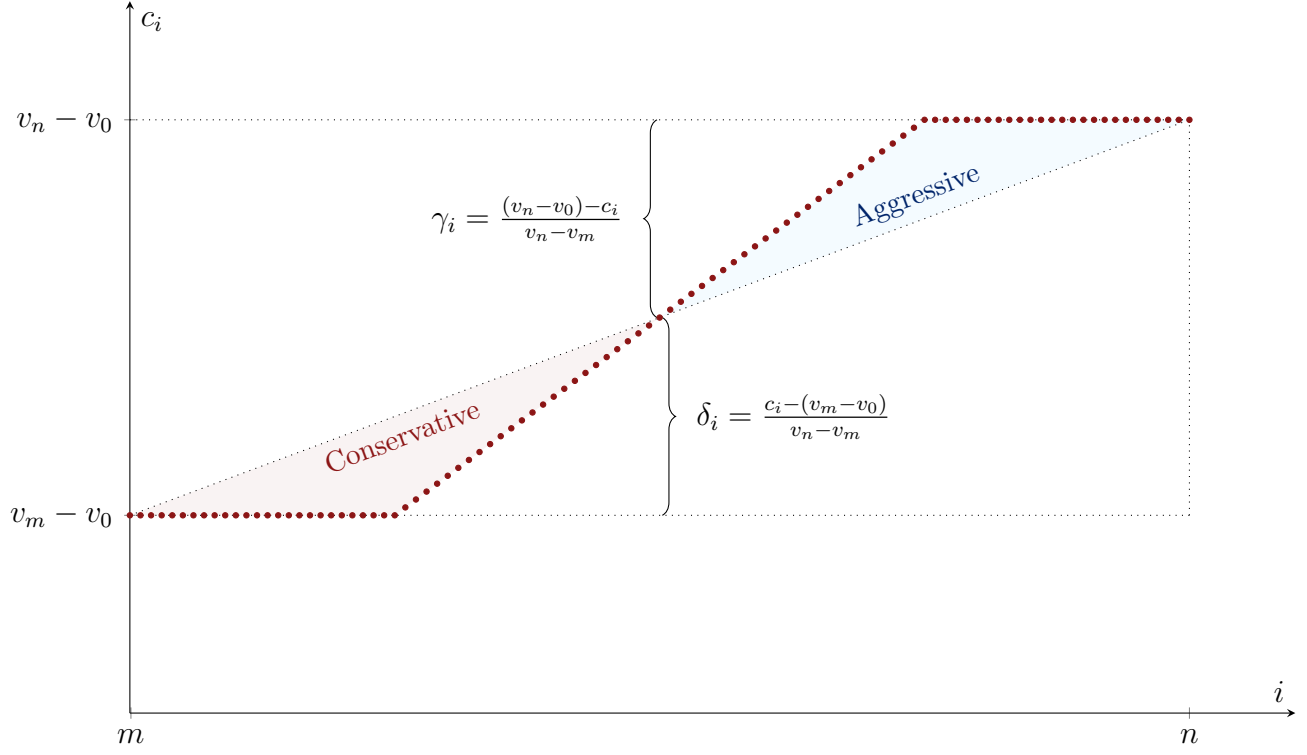


Figure 1.10. Second and third best marginal costs and fundamental earnings distributions with $c_{ii}p_i$ linear in c_i and two informationally distinct ranges ($J = 2$).

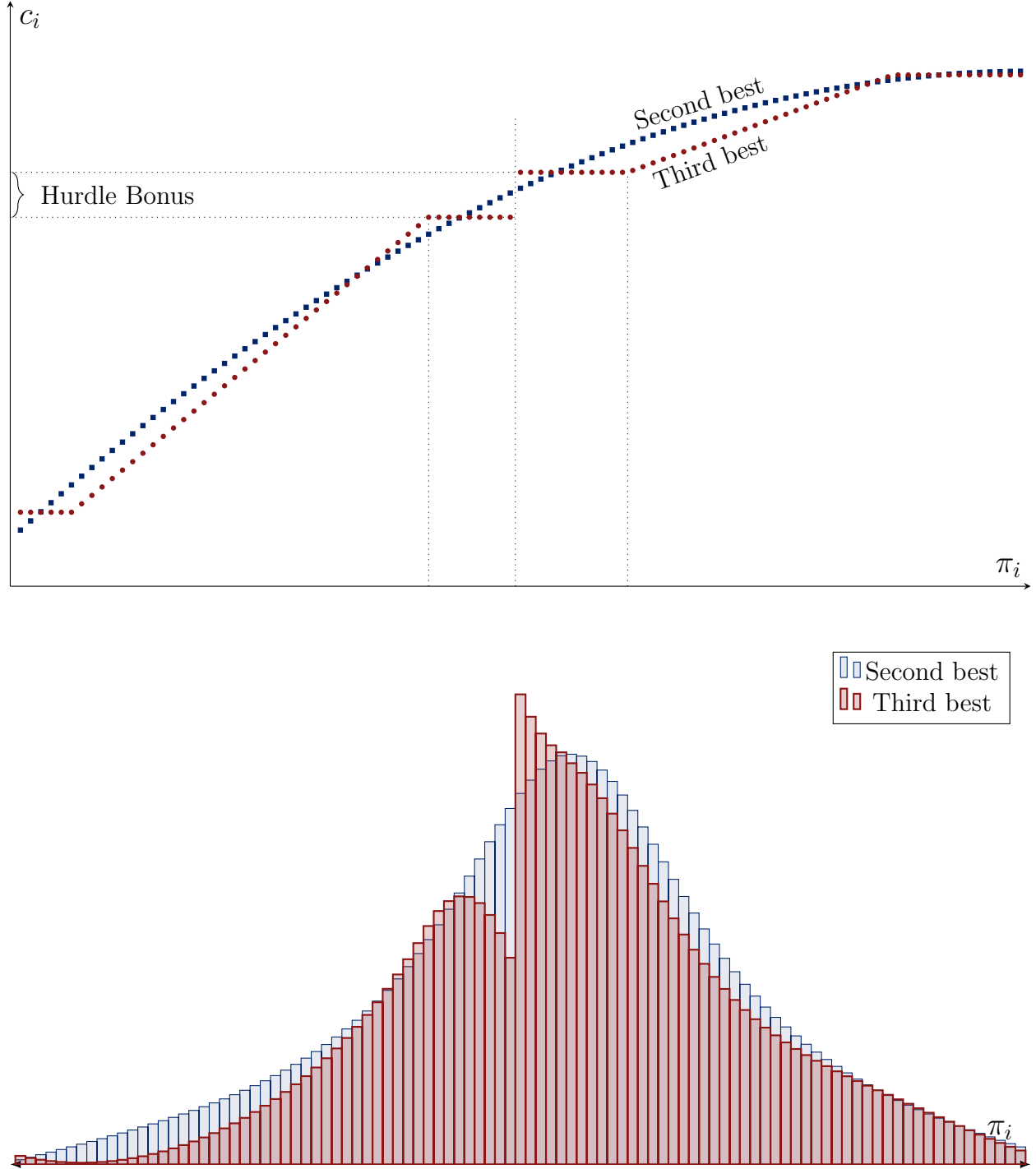


Figure 2.1. The effect of timely loss recognition on downside risk for a firm that prefers concave c_i ; that is, $\phi_i^k + \pi_i$ is increasing concave.

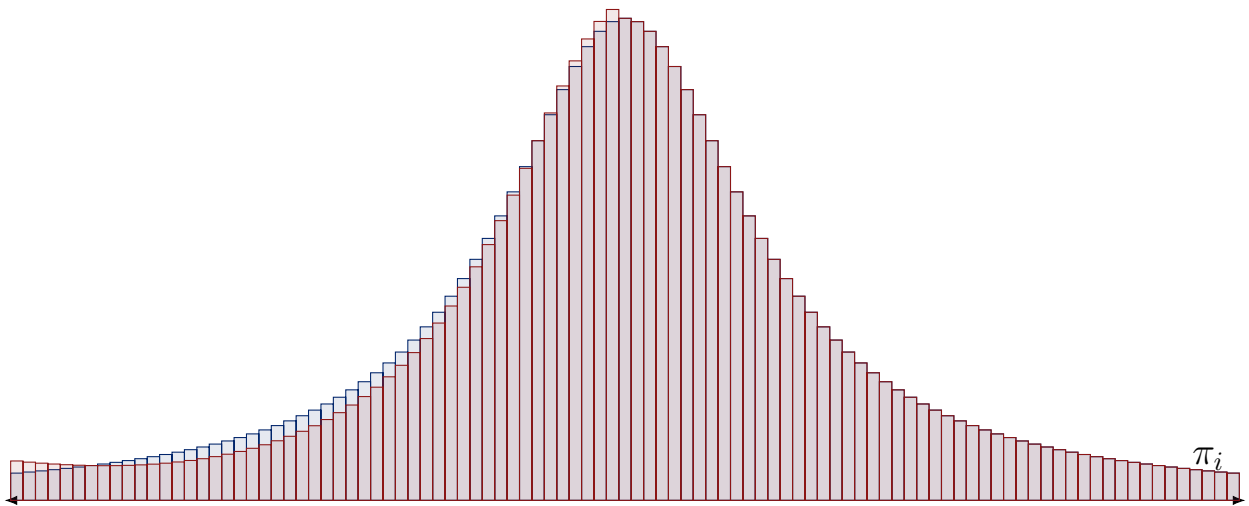
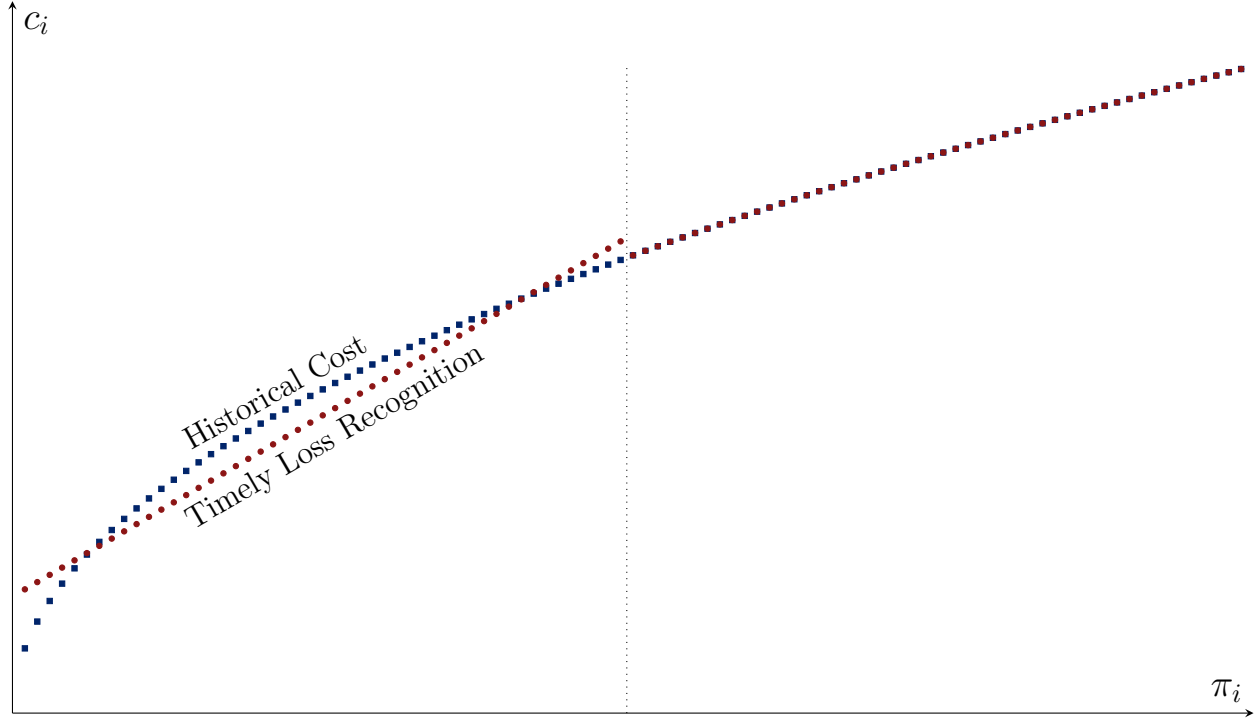


Figure 3.1. Representation of the model with $\delta_i = 0$ for all i .

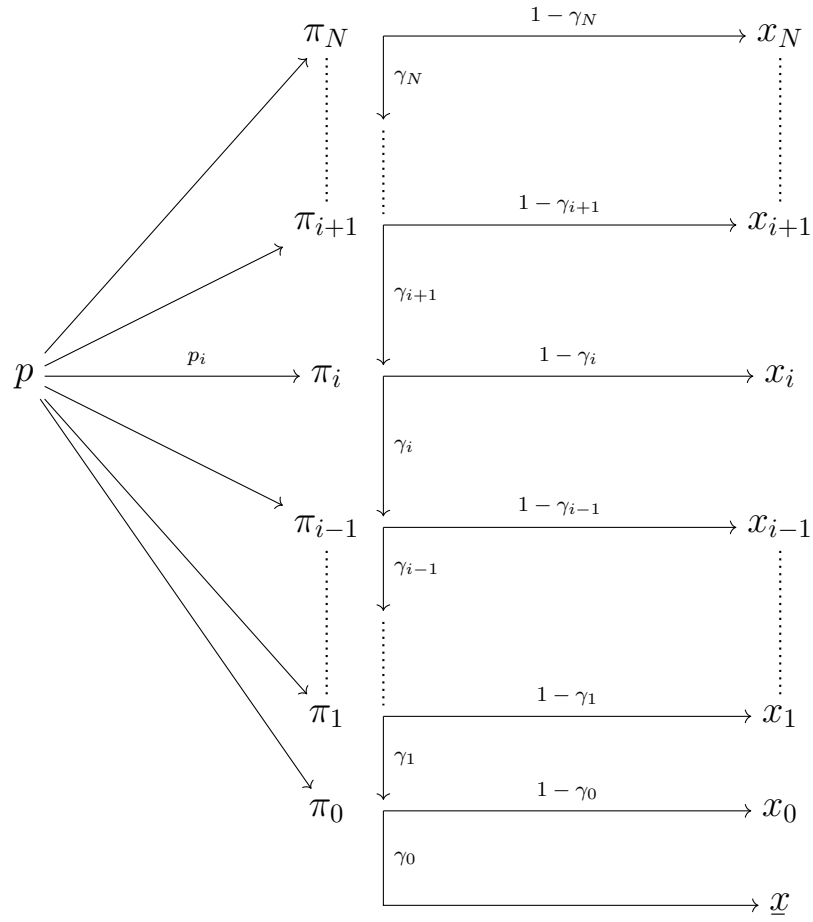


Figure 3.2. RDEU probability weighting functions.